VOLUME XXI NEW SERIES, NO. 155

JOURNAL
OF THE
AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER, 1926

CONTENTS

PAST AND FUTURE DEVELOPMENT OF VITAL STATISTICS IN THE UNITED STATES 257
I. JOHN SHAW BILLINGS AND FEDERAL VITAL STATISTICS. By WALTER F. WILLCOX 257
II. WILLIAM ALEXANDER KING AND THE FEDERAL REGISTRATION SERVICE 1900–1906. By GEORGE H. VAN BUREN 267
III. CRESSY L. WILBUR. By WILLIAM H. GUILFOY 272
IV. 1926 PROGRESS REPORT IN THE CAMPAIGN TO BRING EVERY STATE INTO THE REGISTRATION AREA BEFORE 1930. By WILLIAM H. DAVIS 274
V. PRESENT REGISTRATION SYSTEM. By LOUIS I. DUBLIN 279
AN ITERATION METHOD FOR DETERMINING MULTIPLE CORRELATION CONSTANTS. By TRUMAN L. KELLEY and FRANK S. SALISBURY 282
RELATION OF EARNING POWER TO AGE IN PROFESSIONAL WORKERS UNDER CONDITIONS OF NEARLY FREE COMPETITION. By E. L. THORNDIKE and ELLA WOODYARD 293
DETERMINATION OF CURVILINEAR REGRESSION "SURFACES" IN THE PRESENCE OF OTHER VARIABLES. By MORDECAI EZEKIEL 310
THE TREND-SEASONAL NORMAL IN TIME SERIES. By C. H. WHELDEN, JR. 321
NOTES:
A PARTICULAR CASE OF NON-REPRESENTATIVE SAMPLING. By G. Pietra ... 330
THE SIZE OF FAMILIES. By Elbridge Sibley 333
OCCUPATIONAL GROUPS IN THE UNITED STATES, 1820–1920. By P. K. WHELPTON 335
REVISION OF THE WEEKLY INDEX NUMBER. By Irving Fisher 344
A CRITICISM OF PROFESSOR CHADDOCK'S "PRINCIPLES AND METHODS OF STATISTICS" (347); A BASIS OF FORECASTING NATIONAL INCOME (350); A SCALE FOR PLOTTING LOGARITHMIC CURVES (353); PROGRESS OF THE CENSUS (353); MISCELLANEOUS NOTES (354).
REVIEWS 360
RECENT LITERATURE 375

PUBLISHED QUARTERLY BY THE
AMERICAN STATISTICAL ASSOCIATION
PUBLICATION OFFICE: RUMFORD PRESS, CONCORD, N. H.
EDITORIAL OFFICE: COLUMBIA UNIVERSITY, NEW YORK CITY

Price $1.50 per copy **$5.00 per annum**

THE EMPLOYMENT CLEARING HOUSE
OF THE
AMERICAN STATISTICAL ASSOCIATION

will aid you with your employment problem. It makes no charge to employers for its services. Any member of the Association may register for a position. There is no fee.

The following positions were open in August:

Field	Pay									
	Under $1800		$1800–2400		$2500–3600		$3700–4900		$5000 and over	
	M	F	M	F	M	F	M	F	M	F
Agricultural Economics			1				1			
Business Forecasting					2					
Business—Mfg. and Selling			2		3				1	
Drafting		1								
Economic and Industrial Research			2	1	2		1		1	
Governmental							2	1	1	
Merchandising					1					
Social Work				1			1			
Statistical Secretarial				1						
Technical Computing		1	1				1			

Members were looking for openings as listed:

Field	Pay									
	Under $1800		$1800–2400		$2500–3600		$3700–4900		$5000–and over	
	M	F	M	F	M	F	M	F	M	F
Agricultural Economics	1								1	
Business—Mfg. and Selling			2		3	1	2		5	
Economic and Indus. Research	1	2	1	2	2	3	3		3	
Finance					1					
Insurance						1			1	
Marketing					2					
Social Research				2						1
Teaching			1		1	1	3			
Transportation									1	
Vital Statistics				2	1					

COMMUNICATE YOUR NEEDS TO

HELEN W. LINKE

Secretary's Assistant, American Statistical Association

474 West 24th Street, New York City

[NOT A PAID ADVERTISEMENT]

N

J
S

F
=

NEW SERIES, NO. 155 (VOL. XXI) SEPTEMBER, 1926

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Formerly the Quarterly Publication of the American Statistical Association

THE PAST AND FUTURE DEVELOPMENT OF VITAL STATISTICS IN THE UNITED STATES[1]

I

JOHN SHAW BILLINGS AND FEDERAL VITAL STATISTICS

By Walter F. Willcox, *Cornell University*

John Shaw Billings, we are told by his biographer, achieved distinction in six major fields of activity. With only one of these—vital statistics—the one which many would call the least important of the six, are we primarily concerned tonight. But about another—public health largely brought under tillage by the aid of vital statistics—we shall also have something to say. As Wilbur has written:

> Vital statistics is the Cinderella of modern public health. She sits in the chimney corner and sifts the ashes of dusty figures while her proud sisters bacteriology and preventive medicine go to the ball and talk about the wonderful things they have done. But the Prince's slipper fits no other foot and when we descend to facts about the results of public health work vital statistics are our sole dependence.

Before considering Billings' work in these two fields we must in passing mention the other four. He was great as a designer of hospitals and sanitary engineer; as a medical bibliographer and historian; as an advocate of better medical education and of higher standards in American medicine; and as a civil administrator of ability almost of genius.

In turning to the theme of this paper, the contributions of Billings and his successors to vital statistics and public health, it will not be out of place to speak first of his connection with the two societies which are gathered in joint session to honor his memory and that of the epigoni, King and Wilbur, and to take counsel how we may complete and crown their work.

[1] Papers read at the joint meeting of the American Statistical Association and the American Public Health Association, New York City, April 9, 1926.

The American Statistical Association is one of the oldest statistical societies in the world and is, I believe, the oldest American society devoted to a special field of social study. It was founded in the year after Billings was born and is now, therefore, 87 years old. But through the years of Billings' active life the Statistical Association, although in name a national society, was in reality a local New England association with annual meetings held invariably in Boston. Perhaps in part for this reason Billings was never closely associated with it despite the important work he did for American statistics.

With the American Public Health Association the case is different. That society started when Billings was 34 years old and by his work for the improvement of the Government's marine hospitals which were then in a deplorable state as well as by other achievements he had already made a reputation in the public health field. From the beginning it recruited its membership not by mere application and payment of dues but by selection from those who had an acknowledged interest in or devotion to sanitary studies and allied sciences and to their practical application. Naturally, therefore, when it was launched and even before the constitution was adopted, Billings was elected to membership and naturally, too, he became from the start a working member. Elected in 1872 he attended the annual meeting in New York City the following year; a year later[1] at the meeting in Philadelphia he presented two papers, one on "Hospital Location and Construction," another on "The Effect of Mountain Climates on Health," as indicated by special reports of army medical officers. At the same meeting a committee was appointed with Billings as its chairman "to prepare schedules for the purpose of collecting information with regard to the present condition of public hygiene in the principal towns and cities of the United States and the laws and regulations State and municipal relating to the same." At the Baltimore meeting in 1875 Billings presented a report from this committee which was adopted, and the Executive Committee of which he was made a member was instructed to appoint a special committee to carry out the plan for a sanitary survey of the United States. At the Boston meeting in 1876 Billings read a paper on "The Rights, Duties, and Privileges of the Community in Relation to those of the Individual as Regards Public Health."

At the Richmond meeting in 1878 the all absorbing subject was the epidemics of yellow fever in that and the immediately preceding years. A report was presented from a commission which had been appointed to gather and present the facts. This report was referred to a com-

[1] American Public Health Association *Reports*, Vol. 2, p. 539.

mittee of which Billings was the chairman.[1] In the committee's report[2] after commending the work already done, the chairman urged that the investigation should be continued, and take a much wider range so as to discover if possible the cause of yellow fever. He added that the best place to carry on the study was probably Havana. This was the earliest suggestion, so far as I know, of the commission which was appointed more than twenty years later when Cuba was in American occupation, the commission headed by Walter Reed, to study sanitary conditions in Cuba and especially yellow fever so as to protect the health of the American troops stationed on the island after the Spanish American War.

Between the Richmond meeting of 1878 and the Nashville meeting of 1879 the National Board of Health was created (Act of March 3, 1879) as a result of widespread demand for aid from the Federal Government in dealing with the menace of yellow fever, and Billings was made vice-president. Another initial member of the board, Dr. H. I. Bowditch of Massachusetts, as high an authority in public health work as the country then had, wrote a few years later about Billings' work at Memphis:

> National sanitary power, guided by its efficient agent, Dr. John S. Billings . . . was able to put an absolute *cordon sanitaire* around Memphis and at the same time crush out a terrible epidemic of yellow fever prevailing there which without this external aid from the board would have invaded towns above and below Memphis on the Mississippi River thus causing great destruction of human life as well as much interruption of commerce.[3]

It was no doubt as a result of his work at Memphis in collaboration with the State Board of Health that the informal ballot for president of the American Public Health Association at the Nashville meeting showed two-thirds of all the preferences to be for him.

To bring home to us more vividly the conditions in the southern states before the public health movement began, let me remind you of earlier epidemic conditions in the city of New York. For this purpose I quote from the unprinted reminiscences of an octogenarian New Yorker, recently deceased, about his youth in this city written for the entertainment of his grandchildren.

He described as follows a cholera epidemic which ravaged the city when he was eleven years old:

> There was a fearful panic over the coming of the Asiatic cholera. The city was almost abandoned by such terrified citizens as were able to get away. Grass was seen growing up between the cobble stones in the streets that were usually

[1] *Op. cit.*, Vol. 4, p. 355.
[2] *Ibid.*, p. 361.
[3] *Life and Correspondence of H. I. Bowditch*, 2, 243 *ff.*

> thronged with all kinds of carriages and carts, but which were now almost wholly deserted. I remember walking down Broadway one day about two o'clock in the afternoon and meeting only one or two persons between Canal Street and the City Hall. No stores were open except here and there a druggists or a grocery. Every night the "dead carts" could be heard rumbling over the pavements to gather up the uncoffined bodies of those who had died during the day and dump them into trenches in "Potters Field." Those who were able to give their dead a more decent burial had fled in terror from the city. Very many were said to have died from no disease but simply from fright. . . . The population of the city at that time was about two hundred thousand, and the frightful ravages of the plague may be seen from a single fact. It was a part of my daily duty to go each morning to a drug store on the corner of Broadway and Lispenard street, and learn from a bulletin posted there the number of reported deaths on the day before. And though so many had fled from the city that it seemed almost depopulated, yet one morning, I distinctly remember, the deaths of the previous day that had been made known to the authorities were two hundred and twelve. How many more had for one reason or another failed to be reported, there was no means of discovering.

About those conditions of the city streets and of the water supply which opened the door to the epidemic he wrote:

> The streets of the city were muddy and filthy in the extreme. No provision was made for the disposal of the garbage except by throwing it into the street, where sooner or later it was found and devoured by the hogs that roamed through all the resident part of the city in great numbers and with none to molest or make them afraid, but occasionally a mischievous boy or a persecuting dog. They were recognized as the city scavengers, and many a poor Irishman was practically encouraged to keep as many hogs as he could manage to get, by the fact that it cost him nothing to feed them. I was going along the street one day when I saw an Irishman open the gate of his pig-yard which fronted on the sidewalk, and let out a drove of twenty or thirty hungry hogs to roam about the city and pick up their dinner from the garbage of every kind which the servants had thrown out into the gutter. I was told that they all knew the way home and came back at night well fed.
>
> . . . Our washing water was drawn up by hand from the cistern in the yard, and our drinking water was obtained from the street pumps. These were placed on the edge of the sidewalks every few hundred feet apart. Several times a day the servant would need to go and get a pail of water which would soon become unpalatable, especially in warm weather when fresh and cool water was most desirable, as ice carts and ice delivery had not yet been thought of. This pump water was so far from pure that when the comparatively pure Croton water was introduced, it seemed for a time almost undrinkable so perverted had our taste become by the water to which we had been accustomed. After we had used the Croton water for a time the pump water was absolutely offensive.

But you may say these were conditions before Billings was born. Very true but the tell-tale death-rates indicate that in his early manhood New York conditions were not much better. About twenty years after the above described epidemic of cholera this city was

smitten with a severe epidemic of typhus. The first president of the American Public Health Association was then beginning his public health work. Finding that more than 100 cases of typhus had been received from a single tenement house he inspected it and found it a fever nest. The doors and windows were broken, the cellar was full of sewage and every room above was occupied by families of recent immigrants sleeping upon straw on the floor. It was a favorite refuge for new arrivals because no one was in charge and there was no rent to pay. Conditions like these led to the enactment by the New York legislature of the Metropolitan Health Law modelled upon the English Sanitary Act. A few days after the Metropolitan Health Board was organized a case of cholera in the city was reported. The Disinfecting Department at once took charge of the premises, the patient recovered, and no more cases occurred. Thus for the first time in any part of the United States an epidemic of cholera had been effectively stopped by public health measures.

Let us turn now to the special subject of this evening's meeting, American progress in vital statistics since 1880 or, as I would prefer to say, since 1872 when the American Public Health Association was founded.

First let us ask, What is vital statistics? Here we may take Billings as our authority:

> Vital Statistics in the widest sense of the term includes the records of all circumstances affecting the production or duration of human life, corresponding almost precisely with the term "démographie" as used by Guillard and other modern French writers. It includes records of the population living at a given period such as are obtained by the census; and also a record of the changes taking place in this population by births, marriages and deaths, such as is obtained by registration.

Next we may ask, What was Billings' statistical program? It is nowhere clearly and fully stated but we may reconstruct it from different passages. In a paper read before the Fourth American Meeting of the American Public Health Association in 1876 he wrote: Among the difficulties in the way of positive and effective action to improve public health is "our want of a means of testing the real value and usefulness of the various measures of preventive medicine—which means must be a registration of disease and of mortality; and for providing this very careful and well-considered legislation is necessary."

Four years later in his presidential address before the American Public Health Association he said: "Slowly and steadily there has arisen and is growing a belief that much of our sickness and death is preventable; that we ought to be able to make our cities as healthy

as the country; and from peoples of all conditions . . . comes a demand to put away these plagues which consume our children." Two years later he wrote: "The more familiar I become with the results of registration in the United States, the more I am confirmed in my previous conclusions that the General Government should take the matter in hand and through a permanent census bureau or through the National Board of Health secure by paying for them uniform and complete registration returns from the several States."

This goal which Billings set before himself and the country has not yet been reached but Billings was apparently the first man to have a vision of the goal, and surely the first to take steps towards it and inspire others to work in his spirit and with his restrained enthusiasm to that distant end.

Let me now briefly summarize some of Billings' outstanding contributions to vital statistics.

MECHANICAL METHODS OF TABULATION

While the returns of the Tenth Census were being tabulated at Washington, Billings was walking with a companion through the office in which hundreds of clerks were engaged in laboriously transferring items of information from the schedules to the record sheets by the slow and heart-breaking method of hand tallying. As they were watching the clerks he said to his companion, "There ought to be some mechanical way of doing this job, something on the principle of the Jacquard loom, perhaps, whereby holes on a card regulate the pattern to be woven."[1] The seed fell on good ground. His companion was a talented young engineer in the office who first convinced himself that the idea was practicable and then that Billings had no desire to claim or use it. Thereafter he devoted the bulk of his life with great ultimate profit for himself and the world to ripening the invention and securing its adoption. I have no need to describe or eulogize the Hollerith machines. But it is appropriate to note that the most important mechanical or technical contribution to the development of modern statistics ever made anywhere resulted from incubating an idea thrown off by Billings in what seemed an idle moment.

CORRECTED DEATH-RATES

The death-rate under one year of age is about 50 times as great and that at 90 years of age about 100 times as great as in the healthiest

[1] This statement is based on my memory of a conversation with Mr. Hollerith. I have since received a letter from the daughter of Dr. Billings, Mrs. K. B. Wilson, in which she writes: "I do not remember hearing of Father's remarks to Hermann Hollerith about these machines being applied to census tabulation, but I do remember the first little wooden model which Hermann Hollerith brought to our library many evenings while they were puzzling their brains over its adaptation—my Father had no mechanical gifts—so the entire credit is Mr. Hollerith's."

period of life. Groups of persons between 2 and 53 years of age usually have a death-rate below the average but children under 2 and elderly persons above 53 years of age have a death-rate above the average. The former may be called the healthy and the latter the unhealthy ages. In comparing the death-rates of different communities or classes it is important to know whether the proportions of persons belonging to the various age periods are about the same, otherwise the differences in age composition of the populations may partly or completely mask the real differences in mortality. Since country districts frequently have a larger proportion of children and almost always a larger proportion of aged persons than cities, a comparison between the healthfulness of city and country districts on the basis of their crude rates is very likely to be misleading.

Between 1880 and 1895 several European statists devised a method of meeting this difficulty by the computation of corrected, or standardized, or adjusted rates. The method was first made known in this country in the summer of 1893 when Körösi of Budapest read a paper before the Chicago meeting of the International Statistical Institute on "An International Mortality Standard or Mortality Index." Körösi's paper was printed a few months later by the American Statistical Association. Within less than a year the able Secretary of the Massachusetts State Board of Health and Billings had applied the new method, the former to Massachusetts, the latter to each of the 28 American cities having at least 100,000 inhabitants. King introduced a section on Corrected Death-Rates into the Twelfth Census volumes on vital statistics and eleven years later Wilbur returned to the topic in the annual volumes on Mortality Statistics so that the method which started in this country with Billings and Abbott has now become standard practice. The former wisely added: "A correction of the gross death-rate for peculiarities of race distribution of the population of each city would be much more important than the correction for age distribution." American practice has not yet adopted, as in my opinion it ought to adopt, this suggestion. In other words it has imitated English procedure although Billings saw that the diversities in the social composition of our population demanded the modification and improvement of that procedure.

LIFE TABLES

Corrected or adjusted rates as a measure of health stand between the less accurate ordinary or crude death-rates and the more accurate but far more complicated life table.

Prior to Billings' connection with the Federal census that office had not prepared any significant life tables. To be sure one had been computed at Walker's request by the chief clerk of the Bureau of Statistics in the Treasury Department and included in the Census of 1870. But it was based upon "the partial and fragmentary returns of mortality" in that Census supplemented by what Elliott called "a somewhat arbitrary assumption" namely that "the deficiency in the returns of deaths was forty-one per cent of the full number of deaths which must have taken place." Apparently that number of unreported deaths must have taken place because that number was needed to give a rate of mortality for the United States in substantial agreement with the rates prevailing in England and in Massachusetts. A life table prepared on that assumption was insignificant, or worse, misleading.

It was the great merit of Billings that he was not interested in statistical will-o-the-wisps and discarded all such fantastic statistical computations. Instead he must have reasoned there are two states, Massachusetts and New Jersey, and twelve cities including New York and Brooklyn in which the local records of deaths for a series of years are trustworthy enough to give a sound basis for life tables. So he prepared life tables for each of these districts.

His judgment about the value of life tables for a sanitarian is given in the following quotations: "In order to compare the mortality of one place with that of another or of the same place at different times, in order to estimate the influence which sanitary works . . . have had on the public health, it is desirable to reduce the figures to a common denominator as it were by calculating what they would be if the population were stationary; in other words by the construction of a life table." Two years later he wrote: "The most satisfactory method of comparing the mortality of different localities, periods of time, races, occupations, etc., is by the construction of what are known as life tables."

Billings was the first American to construct reasonably accurate life tables of unselected populations and to use them intelligently as measures of community health. To indicate the value of his work in this field let me deduce from it the answer to a question in which we are all interested: How has the average length of life in New York City changed since 1880? In 1880 the average length of life of whites in New York City, his table shows, was 35.0 years; in Brooklyn 38.6 years. The average of these two weighted by the then population of the cities is 36.2 years. But with this must be combined the average length of life of the colored population, unknown for New York but

known for Washington where it was nearly three-fifths of that of the whites. If the same ratio held in New York-Brooklyn the average duration of life of the colored population in those cities in 1879–80 was 21.6 years and the average duration of life for both sexes and all races was 36.1 years.

Three years ago the Bureau of the Census, continuing and extending this pioneer work of Billings, published some abridged life tables for 1919–20. These showed that the average duration of life among whites in New York City was 53.2 years. The average length of life of the colored in New York City for that year, also, is unknown but in the large cities of the country as a group it was 74.1 per cent of that of the whites in the same cities. If that relation held in New York City, the average length of life of the colored in this city in 1919–20 was 39.4 years and of all races it was 52.8 years.

We now can answer the main question. The average length of human life in New York-Brooklyn in 1879–80 as shown by the Billings' life tables was 36.1 years; that in the present New York City in 1919–20 as shown by the Census Bureau life table was 52.8 years. Thus in forty years the length of life in New York has increased by 16.7 years or 46 per cent.

SANITARY DISTRICTS

We are only beginning to realize the many subtle ways in which geographical conditions affect morbidity and mortality. This change was foreseen by Billings. Earlier censuses had made the state the local unit. But state lines seldom coincide with topographic or climatic lines of division and therefore the state cannot be a satisfactory unit in studying the geographic distribution of disease and death. Counties, on the other hand, are too numerous and too small to serve the purpose. So Billings with the help of the geographer of the census combined those counties within a state which were topographically similar into tabulation units or state groups and those state groups which were topographically similar into grand groups. This method was employed in the Censuses of 1880 and 1890 and in the latter it was extended by tabulating mortality statistics of cities by sanitary districts, the boundaries of which were determined under the advice of the local police and health authorities and with reference to topography, character of habitation and of residents. Since 1890 this method of tabulating and interpreting mortality statistics has been abandoned by the Bureau of the Census. I question whether that change was a wise one. Whether it was or not the developments

in New York City of recent years in that direction might well have taken as their motto "Back to Billings."

DIMINISHING BIRTH-RATE

Billings was the first American scholar to call attention on the basis of nation-wide statistics to the decrease in the American birth-rate. His article was printed in 1893 in *The Forum* of which Walter Page was then editor. It attracted far less attention than it deserved partly because neither it nor its substance was included in the Reports of the Eleventh Census and partly because, both in the Census Reports which were printed after Billings in some indignation had severed his connection with the Census Office[1] and also in the *Political Science Quarterly* of the following year, King gave the evidence which led him to reject Billings' results and conclude that the American birth-rate did not decrease between 1880 and 1890. We know now that Billings was nearer the truth than King but that the figures on which Billings relied somewhat exaggerated the decrease in the decade which he studied. We know, also, that the decrease of the American birth-rate began as early as 1810 and has continued for over a century with an interruption in only one decennial period, that between 1850 and 1860. It is in this field that my own work has built most directly on that of Billings and my grappling with his mind has left in me an admiration almost unqualified for the sanity, range, and penetration of his judgment in dealing with a field relatively unfamiliar to him.

When the International Congress of Hygiene and Demography met at Washington in 1912 Billings, then in his seventy-fifth year, was one of the two American vice-presidents of its demographic section. At the opening meeting of that section its president said: "In the field of demography a foreigner is unlikely to visit the United States, unless drawn hither by curiosity, philanthropy or duty. The sluggish and retarded development of American demography is familiar to all students of the subject. But a new life has been stirring since the twentieth century began."

How much Billings had contributed to that new life I have sought to show. It is summed up in the telegram which Dr. Wilbur, Secretary of the section, sent to Billings: "Greetings and sympathy from opening session to one whose work in demography marks an epoch in the American development of that subject."

[1] "The whole of my work in the census (*sc.* the census of 1890) has been done in the face of great obstacles owing to repeated changes of clerks for political reasons, etc., and I am tired of struggling with the most unpropitious circumstances which have surrounded the work." Quoted from Billings in W. F. Willcox, *Area and Population of the United States at the 11th Census*, in American Economic Association Economic Studies II: 240 (August 1897).

II

WILLIAM ALEXANDER KING AND THE FEDERAL REGISTRATION SERVICE 1900–1906

By George H. Van Buren, *Supervisor, Statistical Bureau, Metropolitan Life Insurance Company*

There appears in the "Letter of Transmittal" accompanying the Census Bureau's Report on Vital Statistics for the year 1880 the following significant sentence.

> The greater part of the tables relating to deaths were prepared under the immediate superintendence of Mr. William A. King, to whom great credit is due for the energy and zeal which he displayed in the work, and for making them as accurate and complete as the data would permit.

In the "Letter of Transmittal" accompanying the "Report on Vital and Social Statistics" of the Census of 1890 appears the following from Carrol D. Wright, Commissioner of Labor, in charge of the Eleventh Census:

> The Report on Vital and Social Statistics was prepared by Dr. John S. Billings, Deputy Surgeon General, U. S. Army (retired), who, in submitting the report, desires to call attention to the valuable aid in its preparation which has been given by Mr. W. A. King, Chief of the Division of Vital Statistics, from the beginning to the completion of the work, and for many suggestions as to method of compilation and modes of presentation of the results.

In the "Letter of Transmittal" accompanying the Census Bureau's "Mortality Statistics, 1900 to 1904" appears the following acknowledgment written one month subsequent to Mr. King's death, by the Honorable S. N. D. North, then Director of the Census.

> Mr. King was the originator of the Standard Certificate of Death now in use in practically all the Registration states and cities; and his untiring zeal in the great work of standardizing vital statistics, and in extending the Registration Area, entitles him to rank among the foremost of the comparatively few men in this country who have devoted their lives to the development of this important branch of statistical science.

In addition to these brief acknowledgments, there are in the literature only a few scattered and indeed desultory references to King or to his work. I have known this for years, but in assembling the data needed for this present brief appreciation of the man, I have verified it beyond all doubt. There is today, for example, absolutely nothing in print descriptive of Mr. King's most outstanding contribution to our vital statistics; for King, the first Chief Statistician for Vital Statistics, was the upbuilder of that splendid office methodology which laid the foundation for the excellent annual mortality reports published since 1900 by the permanent Census Bureau. It is safe to state

that seldom, if ever, has any man made such large, such vital, contributions to any branch of science as did King to American vital statistics, and received so little public recognition or appreciation.

For all this "there is a reason." This man did nothing ostentatiously. Inclined to reticence, handicapped by partial deafness, he shunned public speaking—although he spoke well. And indeed he wrote well—always briefly and succinctly, covering his subject clearly and understandingly. No intelligent person who read King's texts ever failed to get his meaning. No one who wrote to the Census Bureau, and whose letter was referred to King for reply, but whose questions were answered clearly, concisely. His texts, nevertheless, seldom bore his signature when published; and his letters, for the most part, were dictated for the signature of a higher official. His office associates knew the man and gave him due credit; but outside, relatively few persons knew Mr. King or appreciated his contributions to statistical science.

Fully realizing, therefore, the inadequacy of the tribute I am now trying to give him, I shall render it with the assurance that it constitutes a fuller public recognition of Mr. King's ability than has been given heretofore.

William Alexander King contributed four great services to American Vital Statistics.

First—He was the author of the Standard Certificate of Death. Such a certificate was sorely needed in gathering the basic data for mortality statistics. Beginning with the mortality report for the calendar year 1900 the Census Bureau discarded, once for all, the use of mortality data gathered by enumerators, and based its reports solely on data received from registration offices. This, of course, was a great forward step. But even then the opportunity for greater progress toward uniform and comparable statistics was clearly seen by Mr. King. No two states used the same form of death certificate. King, accordingly, drafted such a form; and it is in use today, revised and improved somewhat as to minor details, by nearly all the states in the Union. It lends itself readily to bringing out statements of basic facts much more fully than they could be obtained prior to its use. The standard certificate was Mr. King's greatest contribution toward obtaining better data for tabulation.

Second—He was an important co-worker in drafting what is known as the "Model Law for Vital Statistics." Time does not permit presentation in detail of the content of this law. There was needed, in concrete form, for presentation to the state legislatures, an exhibit of the fundamentals of the legislation required for states to measure

up to the Census Bureau standards for admission to the Registration Area. There had been a great deal of false motion by the most sincere friends of vital statistics in different states. Worthless laws had been put upon the statutes. It was clearly up to the Census Bureau to lay before the legislators just what provisions such a law must incorporate. The "Model Law" was drafted and it has stood the test as a directive influence. Committees which handled the matter in the legislatures were enabled to report favorably on this measure, backed as it was by the authority of the Federal Government, the American Bar Association, the American Public Health Association and other learned societies. The result was the passing of this law by many states, their consequent admission to the Registration Area years in advance of the time this could otherwise have been brought about, and the steady expansion of the Area. King's valuable suggestions in connection with the "Model Law" were thus an important force in the growth of the Registration Area, although most of this expansion took place subsequent to his death.

Third—Mr. King was the prime mover in getting through Congress the first important FEDERAL recognition of the value of statistics of births and deaths as essentials to the progress of medical and sanitary science in preventing and restricting disease and in devising and applying remedial agencies. This Federal recognition took form in the adoption by Congress of a joint resolution, which Mr. King drafted, requesting state authorities to coöperate with the Census Office in securing a uniform system of registering births and deaths. And this Federal endorsement has been one of the most effective instruments in bringing about the needed legislation in the several states. It thus constitutes Mr. King's second contribution to the expansion of the Registration Area.

Fourth—We come now to King's greatest service to American vital statistics: The upbuilding and perfecting of the organization and splendid methodology of the Vital Statistics Division for its work as a *permanent* office. It was in this work that my own association with him was closest; for he chose me as one of his group heads and official confidants. In that capacity I learned to know his wonderful organizing genius, his grasp of details, his power of concentration, his veritable passion for accuracy, his clear-sightedness, his tireless energy. He originated and established in the Census Division of Vital Statistics a wonderful system of checks and cross-checks—a system which made it almost impossible for errors to creep into his published tables—a system which his successors have kept in force with changes, of course, to meet changes in the form of some of the tables. The Census vital

statistics volumes, during and since King's time, are recognized as marvels of accuracy. Since his death they have been expanded and improved by three chief statisticians whose entire competency we all recognize. But the reports of today, as well as those which Mr. King actively supervised, continue to reflect his genius, and stand, to my mind, as testimonials not only to the competency of his successors, but to the excellence of the methods Mr. King devised.

THE FEDERAL REGISTRATION SERVICE, 1900 TO 1906

The period of Mr. King's service as chief statistician in the Permanent Census Office was featured by the material expansion of the Registration Area for deaths. Six states were added: Indiana (for the calendar year 1900), California, Colorado, Maryland, Pennsylvania and South Dakota, in 1906, the year of Mr. King's death.

Efforts to enhance the value of the annual mortality statistics under his administration were determined and systematic. He recognized the great desirability of uniformity in the laws and in the systems and methods relating to the collection and tabulation of mortality data. His term of office was marked by a well-considered movement for uniform registration laws.

It was under Mr. King's régime that the International Classification of Causes of Death was adopted by the Census Bureau. In the establishment of this entirely new method of classifying causes of death, he called to his aid Dr. Cressy L. Wilbur, afterward his successor as chief statistician. The first American edition of the Manual of the International Classification of Causes of Death was printed and distributed by the Census Office. A pamphlet on the "Relation of Physicians to Mortality Statistics," acquainting them with the scope and requirements of the International Classification and emphasizing the necessity of lucidity and accuracy in statements of causes of death was prepared and generally distributed to the medical profession. Physicians were told in this pamphlet that the value of mortality statistics was, largely, their responsibility, as it was impossible to compile accurate statistics without accurate basic data. Physicians were urged not to make the writing of a death certificate a mere perfunctory matter, and were told that such a certificate was not only a legal, but a moral obligation to their patients for the protection of the rights and interests of those dependent on them; and furthermore, that it was a professional obligation of medical men to further the advancement of medical and sanitary science by recording accurately the causes of death of their patients. A pamphlet on "Medical Education in Vital Statistics" was sent to all medical schools. This

explained the movement for the extension of the Registration Area and the importance of mortality statistics, and suggested that medical colleges give courses of instruction on registration methods and vital statistics.

Personally Mr. King was a most likable man, a devoted and beloved husband and father, a splendid companion to his few intimates. His hobby was amateur photography and he did beautiful work. He was an adept at three-cushion billiards and an enthusiastic baseball fan. He carried his statistical training into these recreations. He had a way of charting many billiard shots, and he had his own code for keeping a baseball score, with a symbol for every possible play. He would jot down these symbols with lightning-like rapidity and after the game, would have its full history readily translatable from his coded score card.

In his contacts with his office staff, he was seldom severe, and was generous to a fault. He had a keen sense of humor, and a way of using it in office discipline in lieu of severer measures. I well remember one Monday morning when the young man, who was his personal stenographer, reported for work fortified only by a shave and a hasty breakfast after a continuous session at poker which had started the preceding Saturday afternoon. As might be expected, he simply could not function. Mr. King was immediately "wise" to the situation and diagnosed the cause. He turned to the man and said "Go, son, sleep; and sin no more." That stenographer went, and never, to my knowledge, gave further cause for dissatisfaction. This same man was a speed-marvel as a typist. On one occasion he asked King to excuse him for the afternoon in order that he might go to the races. Mr. King had a rush job on hand and told him that he could not go unless he could type, beforehand, a formidable lot of dictation. The thing appeared impossible; but the stenographer went at it with zeal. His typewriter clicked off the pages at a speed which suggested that it must be electrically driven. Mr. King was actually "rooting" for that boy to win. During the lunch hour the man continued to grind. King beckoned to a few of us. "The races," he explained with a sidewise movement of the head toward the man. That stenographer "made the grade," and as he darted out on a run in time to make the last race train, Mr. King's face bore a satisfied and benevolent grin. That stenographer, incidentally, had the right stuff in him. He took the night course in law at one of Washington's universities, and, some years ago, was elected prosecuting attorney in a western state. And I know that he loved Mr. King.

One of my own first contacts with Mr. King was the result of a small transgression. I had been writing a personal letter during office

hours, about a week after I began work in the Census Bureau. I had made a mistake in addressing the envelope and had then carelessly affixed a stamp. It was while attempting to detach the stamp from the envelope that Mr. King suddenly appeared before me. He reproved me, in fact, but this was the form of his reproof. "Van Buren," he said impatiently, "you can't tear a stamp off an envelope, but you can tear an envelope off a stamp." Suiting his action to his words, he took the envelope from my hands, quickly and neatly detached the stamp, and went away without another word.

I shall always cherish the recollection of my personal, as well as my official association with William Alexander King. If, in my poor way, I have been able to convey to our younger members, in particular, some idea of the magnitude of the debt that American vital statistics owes him, it has indeed been *worth their while* to listen to this tardy, but sincere appreciation of the man.

———

III

CRESSY L. WILBUR, M.D.

By William H. Guilfoy

With the prolonged mental obscuration of Dr. Cressy L. Wilbur, the cause of vital statistics in New York State and in the United States lost the ablest champion of its day.

Cressy Livingston Wilbur was born at Hillsdale, Michigan, March 16, 1865, of Rodney G. and Frances Cressy Wilbur. Graduating from Hillsdale College in 1886 he was the recipient of a Master's Degree in Philosophy within four years. In March, 1890, Bellevue Hospital Medical College of New York City conferred upon him the degree of Doctor in Medicine.

In June, 1891, he married Blanche M. Mead of Hastings, Michigan. He was Chief of the Division of Vital Statistics of the State of Michigan in 1893, succeeding the late Dr. Baker who had earned an enviable reputation in that office. He served as such for 13 years and while holding this position was employed by the Federal Bureau of the Census as Expert Special Agent in charge of extension of the Registration Area of the United States. Incidentally he found time to lecture on vital statistics in the Medical Department of the University of Michigan from 1902 to 1905.

Appointed as Chief of the Bureau of Vital Statistics July 1, 1906, his field of work assumed a national and international scope. He served as delegate of the Bureau of the Census to the International Congress on Tuberculosis at Washington in 1908, being elected at

that meeting Vice-President of Section VI. He was the chief mover in the re-creation of the section on Vital Statistics of the American Public Health Association, carrying his efforts to a successful issue at the meeting of this Association in 1907.

He was official delegate of the United States for the revision of the classification on mortality and morbidity held at Paris, July, 1909, serving as Vice-President and aiding in the preparation of the official list in English.

At the International Congress of Hygiene and Demography held in Washington in 1912, he served as Secretary of the Division of Demography. His membership in the various associations reflected his intense devotion to the forwarding and improvement of vital statistics in this country.

He was a worthy successor to the late William A. King who, following in the footsteps of Dr. John S. Billings, set a high standard of conscientious and thorough work for the cause of vital statistics. To Dr. Wilbur the credit is due for the preparation of the Model Registration Law, and as a result of his constant efforts twenty states enacted laws providing for its adoption.

His membership for many years in the American Public Health Association was devoted exclusively to the building up of the section on Vital Statistics, ever urging its members to spread the gospel of complete registration of sickness and death and adoption of standard tables of mortality and morbidity.

As a member of the American Statistical Association he was constantly bringing forward for discussion problems of a vital statistical nature.

For his *Study and Prevention of Infant Mortality* he was honored by being elected President of the American Medical Association in 1912.

He also held membership in the International Statistical Institute and was a Fellow of the Royal Statistical Society of England.

From the foregoing it is evident that his lines were cast entirely in statistical waters. His maiden efforts in this direction began with his service in the State of Michigan and for 12 years he edited the monthly and annual reports of the State. He served as Special Agent in charge of the extension of the Registration Area between 1901 and 1906. His more intimate connection later with this work, as Chief Statistician of the Bureau of the Census from 1906 to 1914, resulted in the addition of many states to the Registration Area.

He was an able assistant to William A. King from 1901 to 1906. In 1906 the percentage of the population in the Registration Area was 49 and in 1914, when his services as Chief Statistician of the Bureau of the Census ended, the percentage was 66.

California, Colorado, Maryland, Pennsylvania and South Dakota were added in 1906; Washington and Wisconsin in 1908; Ohio in 1909; Minnesota, Montana, North Carolina and Utah in 1910; Kentucky and Missouri in 1911; Virginia in 1913; Kansas in 1914.

The appointment of a new Director of the Census in 1914, whose knowledge of the workings of the Census Bureau could not be appraised as of a high order of merit, led to differences of opinion between him and his Chief Statistician, the latter stoutly maintaining that there should be no change in the long established procedure of gathering material for the mortality and fecundity reports. The result was that Dr. Wilbur, rather than see the undermining of the splendid structure which he had so laboriously built up, resigned his position.

The late Dr. Hermann M. Biggs welcomed him to more placid waters as Director of the Bureau of Vital Statistics of New York State Department of Health in April 1914, in which position he served for several years with distinction, his efforts being concentrated in obtaining a complete registration of births and deaths in the Empire State.

The severing of his relations with the Bureau of the Census acted as a tremendous shock to his sensitive nature. I remember him distinctly at a meeting of the Sanitary Officers of New York State, reading to Dr. Batt and myself a long indictment of the Federal official who forced him to retire.

In my mind the outstanding results in his long and honorable career were the excellent work in producing and supervising the monthly bulletins and annual reports of the State Department at Michigan, his successful work in the extension of the Registration Area of the United States, his aid to William A. King in the preparation of the first Manual of the Classification of the Causes of Death, the editing of the Revised Manual based upon the findings of the International Congress in 1909, the harmonizing of the latter with the Classification of the Royal College of Physicans of England and that of Bellevue and Allied Hospitals.

May I also add a personal tribute to his modesty, his kindly good nature and steadfast adherence to the principles of honest fair dealing with all men.

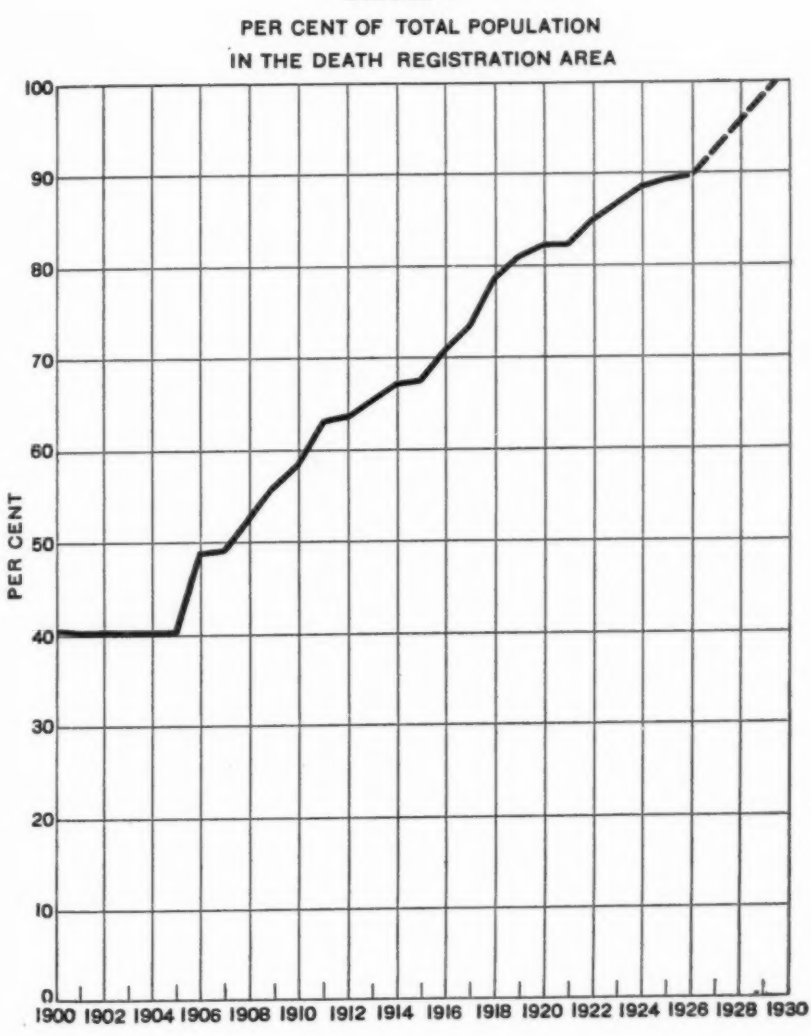
IV

1926 PROGRESS REPORT IN THE CAMPAIGN TO BRING EVERY STATE INTO THE REGISTRATION AREA BEFORE 1930

By William H. Davis, *Chief Statistician, Bureau of the Census*

A year ago I prepared a short paper on "The Necessity for Completing the Registration Area Before 1930." This paper is simply in

FIGURE I

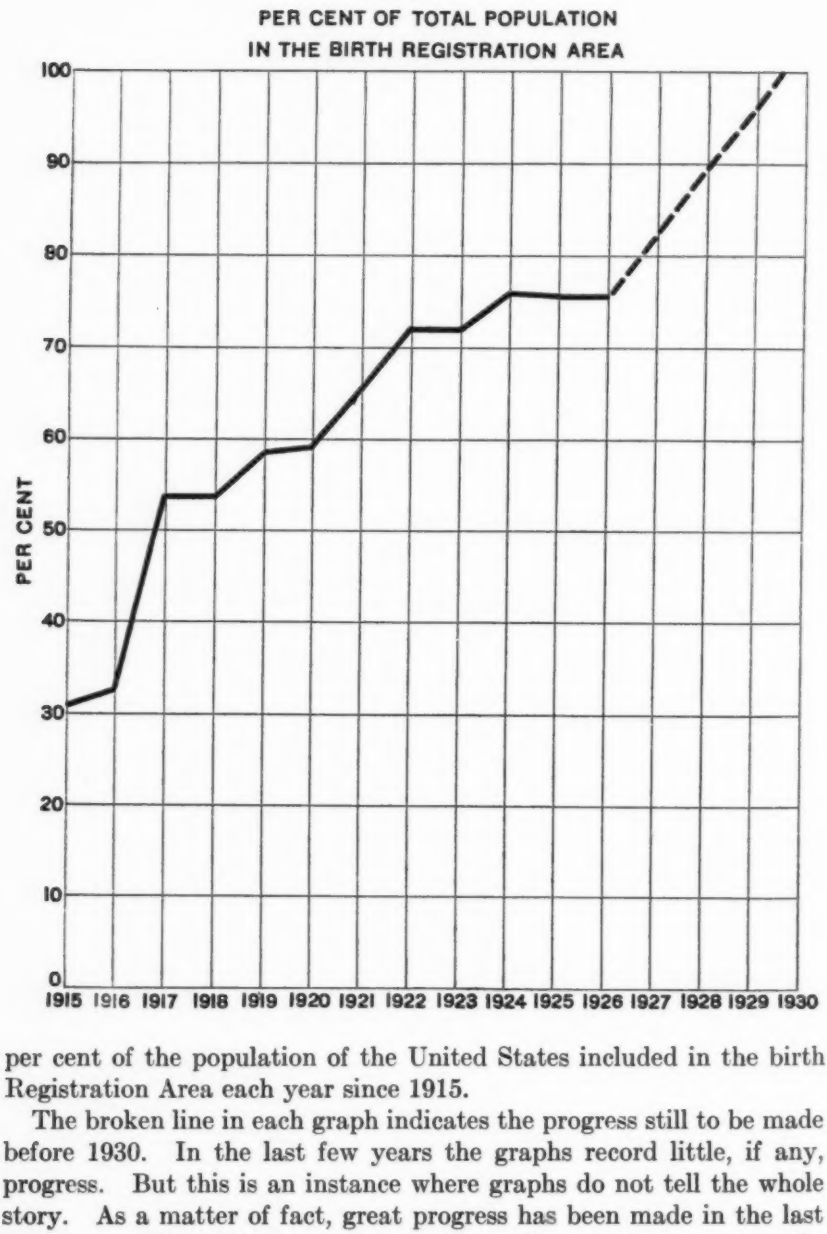


the nature of a Report of Progress in the campaign to reach the goal before 1930.

The two accompanying graphs make clear the growth of these areas in recent years and the growth which must take place in the years 1926, 1927, 1928 and 1929.

In Figure I, the solid black line shows the per cent of the population of the United States included in the death Registration Area each year since 1900, and in Figure II, the solid black line shows the

FIGURE II



per cent of the population of the United States included in the birth Registration Area each year since 1915.

The broken line in each graph indicates the progress still to be made before 1930. In the last few years the graphs record little, if any, progress. But this is an instance where graphs do not tell the whole story. As a matter of fact, great progress has been made in the last few years. There has been no let-up in sowing the seed and already the fields are whitening for the harvest, or if you will permit a different

simile, the growth of this registration movement in recent years has been so rapid that it reminds me of the great big snow balls which small boys roll up when building a snow fort.

Dr. Dublin in his paper tells of the promising start already made by a Committee of the Vital Statistics Section of the American Public Health Association which has as its sole objective the same goal, every state in the Registration Area before 1930, and I will therefore confine my remarks to the recent activities in the same direction by the various states, by the Children's Bureau, by the Bureau of the Census, and by other agencies.

There has been a great registration battle raging in Georgia the past year. Georgia was admitted to the Registration Area for deaths in 1922, but in 1925 the Georgia Supreme Court decided that the registration of births and deaths could not be included under the phrase "necessary sanitation" and that, accordingly, it was unconstitutional for counties to collect taxes for the registration of births and deaths. Without pay, of course, the local registrars could not serve and registration became at once very defective. An attempt to secure a favorable reconsideration by the Supreme Court of Georgia failed. An attempt to secure favorable legislative action at the regular session of the 1925 Assembly also failed. Then the Governor was literally bombarded with appeals to save the day by including this registration matter in his call for the 1926 special session. However, he failed to mention it in his original call and the outlook was gloomy, for unless the question was considered at this 1926 special session, an amendment to the Constitution of the State would have to go over till the regular 1927 session and could not be ratified by the electorate until the fall of 1928. As Georgia now had an unsatisfactory registration law, the State was dropped from the Area. To cut a long story short, Governor Walker a few months ago issued an amended call which included the registration difficulty and the House and Senate have just passed the desired amendment to the Constitution so that the people of Georgia will now have a chance to ratify this amendment in November of this year. By 1927, therefore, we hope to have Georgia back in the fold.

Nevada, South Dakota, and Texas all need better registration laws, and without going into detail, let me say that the outlook is good for favorable action in all of these states at the next session of their respective legislatures in 1927.

I will not attempt to depict the situation in every state. Suffice it to say that the state officials are busy and they are trying to bring their states into the Area.

The Children's Bureau has done and is doing splendid registration

promotion work. The Children's Bureau is primarily interested in accurate birth-rates, accurate infant mortality rates and accurate death-rates from puerperal causes. Without good registration of births and deaths, no such accurate rates are possible. Therefore, the Children's Bureau has considered itself fully justified in using both personnel and funds to improve registration. A rough estimate indicates that at least $20,000 a year of the Shepard-Towner money has been used in this way, to say nothing of the state funds which have been raised to match the Shepard-Towner funds.

The International Health Board has proved and is proving a fairy godmother in improving registration. Its officers have visited many of the states and have given material financial assistance where such assistance was needed.

The Child Health Association has been most generous in loaning one of its agents to states in need. And there have been many other contributors to this campaign.

The Census Bureau form of the Mother's Certificate is now being used in 12 states. Twenty-two other states use their own state forms to promote birth registration.

Posters outlining the reasons for birth and death registration have been placed in the post offices of at least 15 states, and this form of publicity has been suggested to all the states not in the Area.

Four of the regular vital statistics field men of the Bureau of the Census are now working in states which need help, one in Alabama, one in Georgia, one in Mississippi, and one in Texas.

Perhaps the best bit of information is that Dr. Batt, for many years the State Registrar of Pennsylvania, has retired, and it has therefore been possible to secure his services to aid the registration promotion work in the field. Since December 1, he has been a special agent of the Bureau of the Census, devoting his entire time to this work. Drawing from his long experience, he is able to counsel the states wisely and he is sending in most optimistic reports regarding the outlook. He has already visited Tennessee, Arkansas, and Louisiana and thinks that all of these states will be ready for admission to the Area this calendar year.

An extract from a letter received from Dr. Bishop, State Health Commissioner of Tennessee, is illustrative.

> To say that Dr. Batt has been helpful to the Department is putting it very mildly indeed. I have had some ten years' experience in public health work, but I believe I have learned more about the principles of registration of vital statistics since Dr. Batt has been here than in all of my past experience. I believe that with one or two return trips for relatively short stays we will be able to improve our registration at least to the point that will give us entrance

> into the Birth Registration Area, though this measure of improvement will not be our only objective. Without Dr. Batt's assistance I am convinced that we would have been groping in the dark to a certain extent and we would very probably fail to secure the improvement we need so badly. He has gone about reconstruction of our procedure and organization of activity for registration improvement in a practical, systematic way, which I feel will be sure to give results. We are revising our budget with the Children's Bureau, of the Department of Labor, so that his recommendations can be carried into effect. I feel, however, that it is imperative for him to return to Tennessee. . . . *The job is well begun but must be well done.* I have secured a very capable and very well known physician for field work, but this physician must be trained by Dr. Batt.

There are 15 states to be brought into the fold before 1930. Were I asked to specify the year in which each of these states would probably enter the Registration Area, I would predict for 1926 the following five: Alabama, Idaho, Louisiana, Missouri, and Tennessee; for 1927 again five: Arizona, Arkansas, Georgia, New Mexico, and South Carolina; for 1928, three states: Colorado, Nevada, and South Dakota, and for 1929, the remaining two: Oklahoma and Texas.

V

THE PRESENT REGISTRATION SYSTEM

By Louis I. Dublin, *Statistician, Metropolitan Life Insurance Company*

The very unsatisfactory situation as to registration in the United States has resulted from the peculiarities of our Federal Constitution and the division of function between the states on the one hand and the National Government on the other. The framers of the Constitution did not provide for registration of vital statistics; this, therefore, is a state affair. We must not forget that there are forty-eight individual states, each concerned with the problem of registration and each inclined to go its own way, to make its own experiments and sometimes, unfortunately, its mistakes. In view of this condition, much may be said in praise of the Federal authorities who have during the last thirty years labored to bring about a unified and well-organized Registration Area for births and deaths.

Great credit especially belongs to Mr. King under whose guidance the first big step in the organization of the Registration Area for deaths took place and to Dr. Cressy L. Wilbur, who during his incumbency as Chief Statistician, added greatly to the death Registration Area and established the nucleus for the present birth Registration Area. Both of these men had a clear idea of the difficulties in the path. They labored with consummate tact with the state health officers and the registrars, provided them with a good working law and the rudiments of an organization which would function, and kept them encouraged

ever to improve the character of their state and local services. In many respects, having in mind the many difficulties to be overcome, it is an amazing accomplishment that over 92 per cent of the population is now included in the Registration Area for deaths and that 76 per cent of the population is united in a birth Registration Area.

It was at the meeting of the American Public Health Association in St. Louis, October, 1925, that the section on Vital Statistics resolved to do everything in its power to help bring about the completion of the Registration Area by 1930. This resolution called for the appointment of a representative committee which would coöperate with all existing agencies interested in the problem and would coördinate the efforts of the voluntary agencies in support of Dr. Davis and his Division of Vital Statistics of the Census Bureau in the official effort to complete the two Areas. This resulted in the formation of the Committee to Aid Completion of the Registration Area by 1930, of which I am chairman.

Immediately after its organization, the Committee of the American Public Health Association interested the Insurance Committee of the United States Chamber of Commerce in its project. Through the active interest of its Chairman, Mr. Leroy A. Lincoln (General Attorney of the Metropolitan Life Insurance Company), the Committee of the Chamber was able to raise a fund of about $10,000 from a number of insurance companies. This fund was turned over to the American Public Health Association for the assistance of its Committee on Registration. Primarily this fund is to be used to support the efforts of the Census Bureau in its field work by financing such activities as are essential to success but cannot properly be undertaken by the Government itself. Dr. Batt, the Field Agent of the Census Bureau, must have a war-chest on which he can draw. A small monthly sum was therefore put at his disposal for stenographic help, for telegrams and for such other assistance as he would need, but which, on account of certain federal policies, would otherwise be unavailable. It was also the purpose of the Committee to employ a number of additional field agents in such states as would promise good and immediate results. It is planned to have the services of two such agents over a period of four or five months, coöperating with Dr. Batt and concentrating each in a group of adjoining states.

The method which is used in each state where the campaign is on is no longer primarily a problem of legislation. Only in South Dakota, Nevada and Texas are the registration laws defective. In the remaining non-registration states, it is the performance of registration which is at fault in spite of fairly adequate laws. Therefore, the first step is to help state registrars to organize their offices along approved lines.

The machinery of registration is no simple matter and only such a person as Dr. Batt, with his years of experience and his unlimited enthusiasm for this work, could hope to accomplish such results in a brief space of time. The field agents of the Committee will follow suit and carry on in those other states where it would be impossible for Dr. Batt to function because of the limited time. But, equally important is the organization of the voluntary agencies in each non-registration state in support of registration. It is planned to interest groups of citizens who can be counted on to support progressive state administration; such are the chambers of commerce, including the important business men of each community, the federated women's clubs, the local chapters of the American Red Cross, county and state medical societies and especially the public press. A well organized campaign of publicity over a period of several months should result in the education of the public to appreciate the value of good registration and to demand it. With sustained public support, these laws may be enforced and registration brought up to the high standards required by the Census Bureau before a state may be admitted to the Area.

From all accounts, the field work already launched by the Census Bureau with such assistance as it could obtain, is bringing good results. Dr. Davis of the Census Bureau tells me that he counts on the admission of five states to the birth Registration Area in 1926: Alabama, Idaho, Louisiana, Missouri and Tennessee; in 1927, five more, Arizona, Arkansas, Georgia, New Mexico and South Carolina; in 1928, three more, Colorado, Nevada and South Dakota, and in 1929 the remaining two, Oklahoma and Texas. I thoroughly believe that through the additional resources and coöperation which this Committee can make available to the Federal authorities Dr. Davis' claims will be found conservative and that the Area can be completed even earlier than he hopes for.

The completion of the Registration Area for births and deaths is within sight provided all work together. We must bring about a realization throughout the United States of the importance of the campaign. We must remove once and for all the stigma that the great United States is the only first-class nation of the world that cannot even now speak officially for its births and deaths over its entire area. Difficult as it is, we must bring about a condition similar to that prevailing in England and Wales and the other nations of Europe. Our country has the resources, the intelligence, the interest and certainly the need for reaching this goal. We will succeed as we bring these together in a unified and persistent effort. The members of the American Statistical Association and of the American Public Health Association should give their support until the victory is won!

AN ITERATION METHOD FOR DETERMINING MULTIPLE CORRELATION CONSTANTS

By Truman L. Kelley, *Stanford University*, and
Frank S. Salisbury, *Washington State Normal School*

The method here presented utilizes a relationship (equation [16] following) which holds between the correct regression coefficient and the coefficient of correlation with the criterion, or dependent variable. This principle was first used by one of the authors [1] in 1924. It is here proven for the first time. The proof also led to another relationship (equation [1] following) which is fundamental in the present method.

Before explaining these and other formulas used it seems desirable, in view of the fact that certain computers may wish to use this method without studying the theory underlying it, to present first a sample problem carefully explaining each step. This is done in the next few paragraphs and following this the proofs of the various formulas used are given. These latter may be passed over by one content to calculate the regression equation by mere rule of thumb.

Let us start with the data of Table I and Table II, Part 1. Our variables are x_0, x_1, x_2, x_3, x_4, x_5, the latter five being independent variables and the first being the criterion. In Table II, Part 1, .2332 is the correlation between x_1 and x_2, .1986 that between x_1 and x_2^2, etc. The diagonal values 1.000, 1.000, etc., are the correlations between x_1 and x_1, between x_2 and x_2, etc. The first step is to make a preliminary estimate of the regression weights. The better these estimates are, the shorter will be the subsequent process. In the present problem these preliminary estimates are, for the five independent variables, .4, .6, .2, .7, and .4, respectively, as recorded in Table II, Part 2. These preliminary estimates are simply the raw correlations of the variables with the criterion, each kept to the nearest one-tenth decimal place.

Using these weights we multiply the values of Rows 1, 2, 3, 4, 5 of Table II, Part 1, by .4, .6, .2, .7, .4 respectively, obtaining the entries in the body of Table III. The entries in Table III are added by columns, giving the A_p values recorded at the foot of the columns. We thus have $A_1 = .80203$, $A_2 = 1.13018$, etc. We then multiply A_1 by the preliminary weight w_1, A_2 by w_2, etc., and sum, obtaining

[1] Frank S. Salisbury, *A Quantitative Evaluation of Some of the Factors Determining Certain Administrative Procedures in the Public Schools of California*. Thesis, Stanford University, 1924.

TABLE I

CORRELATIONS WITH THE CRITERION

	1	2	3	4	5
r_{0p}.....	.4017	.6003	.2379	.6807	.3553

TABLE II

Part 1: Intercorrelations of Independent Variables						Part 2: Calculation of Regression Weights
	1	2	3	4	5	
1.....	1.0000	.2332	.1986	.2569	.1064	$(.4\ a\ b\ c-.01)d$ $= .197 = \beta_{01\cdot 2345}$
2.....	.2332	1.0000	.1747	.4520	.2139	$(.6\ a\ b\ c+.01)d$ $= .321 = \beta_{02\cdot 1345}$
3.....	.1986	.1747	1.0000	.2628	.0033	$[(.2a-.06)b\ c-.02]d = .023 = \beta_{03\cdot 1245}$
4.....	.2569	.4520	.2628	1.0000	.2989	$(.7\ a\ b+.08)\ c\ d$ $= .441 = \beta_{04\cdot 1235}$
5.....	.1064	.2139	.0033	.2989	1.0000	$(.4\ a\ b-.08)\ c\ d$ $= .130 = \beta_{05\cdot 1234}$

Part 3: Calculation of Successive A_p, σ_e^2, and s Values

Row a..	.41440	.58396	.29428	.64384	.40343	$\sigma_e^2 = .6133402$ $s = .6133402$ $\delta_3 = -.06$
$\delta_{3r_{1p}}$....	−.01192	−.01048	−.06000	−.01577	−.00020	
Row a_1..	.40248	.57348	.23428	.62807	.40323	$\sigma_e^2 = .5816266$ $s = .5990662$ $b = 1.0299842$
Row b..	.41455	.59068	.24130	.64690	.41532	$\sigma_e^2 = .6170287$ $s = .6170287$ $\delta_4 = -.08$
$\delta_{4r_{1p}}$....	−.00851	−.01711	−.00026	−.02391	−.08000	
Row b_1..	.40604	.57357	.24104	.62299	.33532	$\sigma_e^2 = .5569775$ $s = .5886047$ $\delta_4 = .08$
$\delta_{4r_{1p}}$....	.02055	.03616	.02102	.08000	.02391	
Row b_4..	.42659	.60973	.26206	.70299	.35923	$\sigma_e^2 = .6630559$ $s = .6430607$ $c = .9698439$
Row c..	.41373	.59134	.25416	.68179	.34840	$\sigma_e^2 = .6236685$ $s = .6236685$ $\delta_3 = -.02$
$\delta_{3r_{1p}}$....	−.00397	−.00349	−.02000	−.00526	−.00007	
Row c_3..	.40976	.58785	.23416	.67653	.34833	$\sigma_e^2 = .6139021$ $s = .6189105$ $\delta_1 = -.01$
$\delta_{1r_{1p}}$....	−.01	−.00233	−.00199	−.00257	−.00106	
Row c_1..	.39976	.58552	.23217	.67396	.34727	$\sigma_e^2 = .6058069$ $s = .6148935$ $\delta_2 = .01$
$\delta_{2r_{1p}}$....	.00233	.01000	.00175	.00452	.00214	
Row c_2..	.40209	.59552	.23392	.67848	.34941	$\sigma_e^2 = .6176173$ $s = .6208965$ $d = 1.0053094$
Row d..	.40422	.59868	.23516	.68208	.35127	$\sigma_e^2 = .6241931$ $s = .6241931$

$$r_{0\cdot 12345} = r_{3c} = \frac{s}{\sigma_a} = .79006$$

TABLE IV

CALCULATION OF $r_{0\cdot 2345}$

$\delta_{2r_{1p}}$....	−.19700	−.04594	−.03912	−.05061	−.02096	
Row d_1..	.20722	.55274	.19604	.63147	.33031	$\sigma_e^2{}_{-1} = .5037394, s_{-1} = .5450582, e = 1.0820241$
Row e..		.59808	.21212	.68327	.35740	$\sigma_e^2{}_{-1} = .5897661, s_{-1} = .5897661, r_{3(e-1)} = \frac{s_{-1}}{\sigma_{e-1}} =$ .76796

$$r_{01\cdot 2345} = \sqrt{\frac{r^2_{3c} - r^2_{3(e-1)}}{1 - r^2_{3(e-1)}}} = .290$$

TABLE III

CALCULATION OF FIRST A_p, σ_c^2, AND s VALUES

	1	2	3	4	5
1	.40000	.09328	.07944	.10276	.04256
2	.13992	.60000	.10482	.27120	.12834
3	.03972	.03494	.20000	.05256	.00066
4	.17983	.31640	.18396	.70000	.20923
5	.04256	.08556	.00132	.11956	.40000
A_p	.80203	1.13018	.56954	1.24608	.78079

$\sigma_c^2 = 2.297400$
$s = .18705$
$a = .5166928$

2.297400, which is σ_c^2. We also multiply r_{01} by w_1, r_{02} by w_2, etc., and sum, obtaining 1.18705, which is designated s. Dividing s by σ_c^2 we obtain a quantity called a, equal to .5166928. We now indicate the multiplication of each of our original weights, .4, .6, etc., by this quantity a as shown in Table II, Part 2, and we also multiply each A_p by it, obtaining values which are entered in Row a of Table II, Part 3. We also multiply s of Table III by a, obtaining s of Row a, Table II, Part 3, and we multiply σ_c^2 of Table III by a^2 obtaining σ_c^2 of Row a, Table II, Part 3. After first noting that in Table II, Part 3, Row a, σ_c^2 and s are equal we compare each A_p value in row a with the corresponding r_{0p} value in Table I and note that the greatest difference in value is in the case of variable 3. We then calculate, δ_3, to one significant figure only, as further refinement unduly increases the arithmetic labor, by equation [1].

$$\delta_p = \frac{r_{0p} - A_p}{1 - \frac{r_{0p} A_p}{s}} \text{ (to be used only in case } \sigma_c^2 = s) \ldots\ldots\ldots [1].$$

We find $\delta_3 = -.06$. We add this value to the preceding weight of variable 3, obtaining (.2a − .06) as shown in Part 2. The values in Row 3, Part 1, are now multiplied by δ_3 and the products recorded in the row labeled $\delta_3\ r_{3p}$. These values are added to the values immediately preceding in Row a, giving the values of Row $a3$. The σ_c^2 value of this row is obtained from the σ_c^2 immediately above by equation [2].

$$\sigma_c^2 \text{ corrected for } \delta_p = \sigma_c^2 + \delta_p(A_p + [A_p \text{ corrected for } \delta_p]) \ldots\ldots [2].$$

Thus

$$\sigma_c^2 = .6133402 - .06(.29428 + .23428) = .5816266.$$

The value s of this row is obtained from the s immediately above by equation [3].

$$s \text{ (corrected for } \delta_p) = s + \delta_p \, r_{0p} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3].$$

Thus

$$s = .6133402 - .06(.2379) = .5990662.$$

We now note that s and σ_c^2 differ considerably and we therefore calculate a new constant b just as we did a before.

$$b = \frac{s}{\sigma_c^2} = \frac{.5990662}{.5816266} = 1.0299842.$$

We enter b as a multiplier of our preceding weights, obtaining weights $.4ab$, $.6ab$, $(.2a - .06)b$, $.7ab$, $.4ab$, as shown in Part 2. We also multiply s by b, σ_c^2 by b^2, and the A_p values of Row $a3$ by b, obtaining Row b. Again noting that $\sigma_c^2 = s$, we compare the A_p values of Row $a3$ with the corresponding r_{0p} values and find that the greatest discrepancy is in the case of variable 5. We accordingly calculate δ_5 by equation [1] and obtain $-.08$. This value is added to the preceding weight of variable 5, yielding $(.4ab - .08)$ as the new weight. Also the values of Part 1, Row 5, are multiplied by δ_5 and the products recorded in Row $\delta_5 \; r_{5p}$. These values are now added to the immediately preceding values of Row b, yielding Row $b5$. The new σ_c^2 value of this row is obtained from the preceding row by means of Formula [2] and the new s from the preceding s by formula [3].

We might now note that σ_c^2 does not equal s and determine a new constant similar to the constants a and b, but a very hasty comparison of the A_p values with the related r_{0p} values shows that the greatest discrepancy is clearly to be found in the case of variable 4. We will, therefore, not take the time to calculate a new constant similar to a and b and new σ_c^2 and s values, but will immediately make the appropriate correction for this variable. We find δ_4 by formula [4].

$$\delta_p = \frac{\sigma_c^2 \, r_{0p} - A_p \, s}{s - r_{0p} \, A_p} \text{ (to be used when } \sigma_c^2 \text{ does not equal } s) \ldots . [4].$$

This gives

$$\delta_4 = \frac{(.5570)(.6807) - (.6230)(.5886)}{.5886 - (.6807)(.6230)} = .08.$$

We add this value to the earlier weight of x_4, giving $(.7ab + .08)$ and, as before, calculate the values given in the $\delta_4 \; r_{4p}$ Row and add this to the preceding row, giving Row $b4$. The σ_c^2 and s of this row are determined by formulas [2] and [3].

Since σ_c^2 does not equal s and since it is not immediately obvious by a comparison of the values of Row $b4$ with the r_{0p} values, which weight is now in most need of correction, we will calculate

$$c = \frac{s}{\sigma_c^2} = .9698439$$

and multiply all the weights in Part 2 thus far derived by this constant, multiply s by it, multiply σ_c^2 by it squared, and multiply the A_p values of Row $b4$ by it, obtaining Row c. A comparison of the values of this row with the r_{0p} values suggests the correction of the weight for variable 3, which is accordingly done next.

The process is continued for a few steps further, leading to Row d. Now the values of Row d are so nearly the same as the r_{0p} values that the next correction would be one affecting the third decimal place only. If we consider that accuracy to two decimal places is sufficient in the weights we will stop here and call the problem solved, except for the next step which is merely a step to check the numerical accuracy of the preceding work.

The next step should be to build up a new table, such as Table III and check the A_p, σ_c^2, and s values of this new table against those of Row d. If the check is close, numerical accuracy is established and the problem is solved. It only remains to use formula [5]

$$r_{0c} = r_{0 \cdot 12 \ldots n} = \frac{s}{\sigma_c} \qquad [5].$$

to determine the multiple correlation coefficient.

For our problem we find from Row d that this multiple correlation coefficient is equal to .79006.

For the sake of anyone who may wish to carry the approximation further and get a closer check, the following values correct to seven decimal places are given

$$\begin{aligned}
\beta_{01 \cdot 2345} &= .19412341 \\
\beta_{02 \cdot 1345} &= .32392693 \\
\beta_{03 \cdot 1245} &= .02748474 \\
\beta_{04 \cdot 1235} &= .43693997 \\
\beta_{05 \cdot 1234} &= .13466545 \\
r_{0 \cdot 12345} &= .79009053
\end{aligned}$$

Table IV gives the steps in the calculation of $r_{0 \cdot 2345}$ and thus by the use of formula [6], of the partial coefficient of correlation.

$$r_{01 \cdot 23 \ldots n}^2 = \frac{r_{0 \cdot 12 \ldots n}^2 - r_{0 \cdot 23 \ldots n}^2}{1 - r_{0 \cdot 23 \ldots n}^2} \qquad [6].$$

For the present problem we have

$$r_{0^2 1\cdot 2345\ldots n} = \frac{(.79006)^2-(.76796)^2}{1-(.76796)^2} = (.290)^2.$$

Since $r_{01\cdot 2345}$ must be of the same sign as $\beta_{01\cdot 2345}$, or w_1, we have,

$$r_{01\cdot 2345} = .290.$$

We will now turn to the theory underlying this method. It has probably already been surmised that the symbols used have the following meanings.

r_{0p} stands for the correlation between the criterion and the independent variable x_p.

w_p is the weight of x_p.

r_{qp} is the correlation of the independent variable x_q with the independent variable x_p.

After q is assigned a value p takes in turn all the values, including q, of the independent variables. When p takes the value q, we have r_{qq}, which is the correlation between a variable and itself, or 1.00.

If, in the table of intercorrelations between the independent variables (Table II, Part 1) the values in the row for variable x_q are multiplied by w_q and similarly for every other row and if columns are then summed the sums, which for column p we will designate A_p, are as given in the following equation.

$$A_p = \Sigma w_q r_{qp} = w_1 r_{1p} + w_2 r_{2p} + \ldots + w_n\, r_{np} \ldots\ldots\ldots\ldots\ldots\ldots [7].$$

in which one of the terms is $w_p\, r_{pp}$, or simply w_p. If we multiply each A_p by w_p and sum, we obtain the quantity called σ_c^2, thus

$$\sigma_c^2 = \Sigma w_p A_p = w_1 A_1 + w_2 A_2 + \ldots + w_n A_n \ldots\ldots\ldots\ldots\ldots\ldots [8].$$

Similarly, the quantity called s is given by the equation

$$s = \Sigma w_p r_{0p} = w_1 r_{01} + w_2 r_{02} + \ldots + w_n r_{0n} \ldots\ldots\ldots\ldots\ldots\ldots [9].$$

We will now establish several facts: First: let,

$$c = \beta_{01\cdot 23\ldots n}\, z_1 + \beta_{02\cdot 13\ldots n}\, z_2 + \ldots + \beta_{0n\cdot 12\ldots n-1}\, z_n \ldots\ldots\ldots [10].$$

in which the β's are the regression coefficients when variables are expressed in terms of standard measures, or measures of deviation divided by standard deviations, and the z's are such standard measures. With c defined as above, r_{0c} is, by definition, the multiple correlation. Thus we have,

$$r_{0\cdot 12\ldots n} = r_{0c} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [11].$$

Second: We may evaluate r_{0c} in terms of s and σ_c. Since all the z's are measures whose means are zero, and since,

$$\frac{\Sigma z_1^2}{N} = \frac{\Sigma z_2^2}{N} = \ldots = 1.00$$

and since $\Sigma z_1 z_2 = N r_{12}$; $\Sigma z_1 z_3 = N r_{13}$; etc., we have, by definition,

$$\sigma_c^2 = \frac{1}{N}\Sigma(\beta_{01\cdot 23\ldots n}\, z_1 + \beta_{02\cdot 13\ldots n}\, z_2 + \ldots + \beta_{0n\cdot 12\ldots n-1}\, z_n)^2$$

which, in case the weights used are these regression weights, may be written in abridged notation,

$$\begin{aligned}\sigma_c^2 &= \frac{1}{N}\Sigma(w_1 z_1 + w_2 z_2 + \ldots + w_n z_n)^2 \\ &= w_1^2 + w_2^2 + \ldots + w_n^2 + 2w_1 w_2 r_{12} + 2w_1 w_2 r_{13} + \ldots \\ &= \Sigma w_p A_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [12].\end{aligned}$$

as is immediately obvious if $\Sigma w_p A_p$ is expanded. By definition,

$$\begin{aligned}r_{0c} &= \frac{\Sigma z_0(w_1 z_1 + w_2 z_2 + \ldots + w_n z_n)}{N(1)\sigma_c} \\ &= \frac{w_1 r_{01} + w_2 r_{02} + \ldots + w_n r_{0n}}{\sigma_c}\end{aligned}$$

which gives by using relationship [9],

$$r_{0c} = \frac{s}{\sigma_c} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [5].$$

Third: The proof that $r_{0p} = A_p$ when correct regression weights are involved in A_p can be deduced from known relationships expressed in determinantal form. Let Δ be the major correlation determinant as follows:

$$\Delta = \begin{vmatrix} 1 & r_{01} & r_{02} & \ldots & r_{0p} & \ldots & r_{0n} \\ r_{01} & 1 & r_{12} & \ldots & r_{1p} & \ldots & r_{1n} \\ r_{02} & r_{12} & 1 & \ldots & r_{2p} & \ldots & r_{2n} \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ r_{0p} & r_{1p} & r_{2p} & \ldots & 1 & \ldots & r_{pn} \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ r_{0n} & r_{1n} & r_{2n} & \ldots & r_{pn} & \ldots & 1 \end{vmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [13].$$

Then, in terms of the minors of this determinant,

$$w_1 = \frac{(-1)^2 \Delta_{01}}{\Delta_{00}}$$

$$w_2 = \frac{(-1)^3 \Delta_{02}}{\Delta_{00}}$$

. .
. .
. .

$$w_p = \frac{(-1)^{p+1} \Delta_{0p}}{\Delta_{00}}$$

. .
. .
. .

$$w_n = \frac{(-1)^{n+1} \Delta_{0n}}{\Delta_{00}}$$

Let,

$$f = r_{0p} - A_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [14].$$

If the weights used in A_p are regression equation weights then multiplying f by Δ_{00} gives the function Φ,

$$\Phi = \Delta_{00} f = r_{0p}\Delta_{00} - r_{1p}\Delta_{01} + \ldots + (-1)^p \Delta_{0p} + \ldots + (-1)^n r_{pn}\Delta_{0n} \ldots [15].$$

Now [15] differs from [13] only in that the elements of the p row are found in place of the elements of the first row. Thus Φ is simply a determinant in which the first row and the p-th row are identical and is thus equal to zero. We have accordingly proven that $\Phi = 0$ when the weights are regression equation weights. If Φ is zero under these conditions then f is also. Having proven the important proposition that,

$$r_{0p} = A_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [16].$$

when regression equation weights are used, it follows at once that,

$$w_p r_{0p} = w_p A_p$$

and

$$\Sigma w_p r_{0p} = \Sigma w_p A p$$

$$s = \sigma_c^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [17].$$

Equations [16] and [17] provide the key to the present treatment, for we work toward a condition wherein these conditions hold. In brief outline the steps are: First, choose any weights which seem reasonable. Second, using these weights calculate the A_p's, σ_c^2 and s.

Since the correlation is not changed if the weights chosen are multiplied by a constant a our third step is to multiply the weights, the A_p's and s by such a constant a, and σ_c^2 by the same constant squared, as will make $\sigma_c^2 = s$. Obviously this constant is given by the equation,

$$a^2\sigma_c^2 = as$$

or

$$a = \frac{s}{\sigma_c^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [18].$$

Having done this we obtain weights which meet condition [17], but condition [16] is not met for each of the variables separately. Thus, fourth, we find the variable with reference to which condition [16] is least approximated and proceed to change the weight for this variable. The optimum change to make is such as will raise the multiple correlation, r_{0c}, the most. We have,

$$r_{0c} = \frac{s}{\sigma_c^2} = \frac{(s - w_p r_{0p}) + w_p r_{0p}}{(\sigma^2 - 2w_p A_p + w_p^2) + 2w_p[(A_p - w_p) + w_p] - w_p^2} \quad \ldots\ldots [19].$$

in which the three terms in ( ) are devoid of the variable w_p and are thus constants with respect to a change in w_p. That these ( ) terms lack w_p is obvious upon expansion, thus, e. g.,

$$s - w_p r_{0p} = w_1 r_{01} + w_2 r_{02} + \ldots + [w_p r_{0p}] + \ldots + w_n r_{0n} - [w_p r_{0p}]$$

and as the [ ] terms cancel w_p is eliminated from the expression. Taking the first derivative of [19] with respect to w_p and setting it equal to zero will give an optimum new w_p weight. It will, however, be more convenient to determine merely the optimum change in the weight of w_p rather than the new weight itself. We will accordingly let the new weight equal $w_p + \delta_p$ wherein w_p is the old weight and a constant while δ_p is our variable. We then have,

$$r_{0c} = \frac{(s - w_p r_{0p}) + [w_p + \delta_p] r_{0p}}{\left\{(\sigma_c^2 - 2w_p A_p + w_p^2) + 2[w_p + \delta_p][A_p + \delta_p] - [w_p + \delta_p]^2\right\}^{\frac{1}{2}}}$$

$$= \frac{s + \delta_p r_{0p}}{(\sigma^2 + 2\delta\ A_p + \delta^2)^{\frac{1}{2}}}.$$

Taking the first derivative with respect to δ_p and setting it equal to zero gives us,

$$\delta_p = \frac{\sigma_c^2 r_{0p} - A_p\, s}{s - r_{0p} A_p} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4].$$

In case $\sigma_c^2 = s$ this immediately reduces to

$$\delta_p = \frac{r_{0p} - A_p}{1 - \frac{r_{0p}A_p}{s}} \quad \ldots\ldots [1].$$

The optimum change to make in a weight is given us under all conditions by equation [4] and under certain special conditions which we can readily bring about by equation [1].

Since a change in the weight of one variable has an effect upon the weights of the other variables and since the weights of other variables than the one in question will commonly need to be changed there is no advantage commensurate with the labor involved in keeping a δ_p value beyond one significant figure. Accordingly we determine the value of δ_p by [4] or [1] to one significant figure and add this weight to the preceding w_p.

Fifth, we must find the effect of this change in weight upon our other constants and obtain new A's, a new σ_c^2 and a new s. It is only necessary to multiply r_{1p}, r_{2p}, etc. by δ_p and add the products to A_1, A_2, etc., respectively, to obtain the new A's. Also as can easily be proven the new σ_c^2 is given by the equation,

$$\sigma_c^2(\text{after } \delta_p \text{ correction}) = \sigma_c^2 + \delta_p[A_p + (\text{new } A_p)] \quad \ldots\ldots [20].$$

The new s is given by,

$$s(\text{after } \delta_p \text{ correction}) = s + \delta_p r_{0p} \quad \ldots\ldots [21].$$

The new multiple correlation is equal to (new s)/(new σ_c) and is of course greater than the preceding r_{0c}. As every step in this process yields a larger r_{0c} than before, the method is convergent, and that without any retrogressive movements, toward the final regression weights and multiple correlation.

The preceding treatment has not yielded the partial correlation coefficients, $r_{01\cdot23\ldots n}$, $r_{02\cdot13\ldots n}$, etc. Just as, when determinants are employed, a second calculation must be made after the β regression coefficient is obtained in order to secure a partial correlation coefficient, so here additional steps are required if this latter is desired. The steps to be followed are relatively brief because when one variable, say the first, is dropped out of a battery the relative weightings of the remaining $(n-1)$ variables determined involving x_1 will generally be very close approximations to their relative weightings when variable x_1 is not involved in the problem. We have the following relationship between multiple and partial correlations:[1]

$$r_{0^2 1\cdot23\ldots n} = \frac{r_{0^2\cdot12\ldots n} - r_{0^2\cdot23\ldots n}}{1 - r_{0^2\cdot23\ldots n}} \quad \ldots\ldots [6].$$

[1] T. L. Kelley, *Statistical Method*, Formula [270].

The multiple correlation $r_{0\cdot 23\ldots n}$ is to be determined along the same lines as was $r_{0\cdot 12\ldots n}$, with this difference, that we will in all probability have very close approximations to the proper weights for variables $x_2, x_3, \ldots x_n$ if we assume the weights already found for them when x_1 was in the calculation.

Table IV illustrates the procedure in the case of variable 1. The first step is the correction of the weight of variable 1, which is .197, by an amount $\delta_1 = -.197$, thus securing A_p, σ_c^2, and s values which are independent of variable 1. It may happen, as in the present illustrative problem, that further corrections would be trivial so that the quotient of s and σ_c immediately yields $r_{0\cdot 23\ldots n}$. Equation [6] then yields the partial correlation desired.

Having given the proof of the formulas employed, the value of the method rests upon its practicability or simplicity. The illustrative problem given possessed but 6 variables merely because a larger number would take up too much space across a book page. The method has proven very simple and straightforward upon a 10 and a 16 variable problem upon which it has been tried. That it cuts down the labor of a 16 variable problem from that of a solution by determinants by 95 per cent is a conservative estimate.

THE RELATION OF EARNING POWER TO AGE IN PROFESSIONAL WORKERS UNDER CONDITIONS OF NEARLY FREE COMPETITION

By E. L. Thorndike and Ella Woodyard, *Institute of Educational Research, Teachers College, Columbia University*

In connection with a general study of the age changes in intellect and ability to learn in adults, we have gathered certain facts which seem of some value to economics. They concern the salaries of clergymen of the Methodist Episcopal Church who lived to 70 or beyond, from age 30 until 70.

The work of a clergyman in this Church is a mixture of preaching, teaching, conducting various religious ceremonies, managing his church as a social enterprise, and rendering personal service in visits to his parishioners, advice, comfort, and aid in sickness or distress. During the time under consideration (from 1860 to 1924) the educational and managerial work has increased as compared with the preaching and personal visitation. On the whole the Methodist pastorate has been a career in which many of the abilities required for success in law, dramatic art, teaching, business, and medicine are usable.

The work is highly competitive. The Church admits certain men as members of its various sections called "conferences," such as the Maine Conference, the East Maine Conference, etc. Every man so admitted is entitled to a parish or "charge," to which he is appointed annually by the bishop who presides over that conference in that year, until he retires temporarily or permanently. There are arrangements for transfers from one conference to another. There are many parishes or churches or "charges," each supported financially by its local members[1] who are represented by an "official board" which chooses annually the man whom the members wish to have as their pastor and which tries to secure his consent and that of the bishop. The bishop has full power, but is advised by the district superintendents or "presiding elders," each of whom has general supervision over a part of a conference, say fifty churches. A minister may refuse to accept the post which the bishop offers him in any year, retiring temporarily for that year; but this would injure his repute, and is rarely done. A church may not refuse to accept the minister whom the bishop

[1] With assistance occasionally from general missionary funds, supplied by a central body for the entire country. This, however, was exceedingly rare for the churches considered in this study.

appoints, and it would not be decent for it to reduce existing salary provisions for that year, though this is sometimes done. But it can make it very clear to him or to the district superintendent that it will make a reduction the following year if he is reappointed. It can also make promises to the candidate whom it does desire.

There are thus n ministers competing for a better appointment than that which they hold; and n parishes[1] competing for ministers whom they regard as superior to the one they have or to others who might be allotted them. The district superintendents compete to get the best possible men for their several districts, in the reassignment which is made annually. Moreover, until 1900 no individual could stay at any one parish or "charge" for more than three successive years, even if he, the official board of that parish, the district superintendent, and the bishop all desired that he should do so. Since 1900, a minister may legally remain indefinitely, but the turnover has been much the same as before. The average length of stay for 90 parishes taken at random, whose records from 1900 to 1924 were examined, was under three years.

In general, then, salary measures estimated ability rather closely. Any parish (save the few paying the highest salaries) can obtain ministers of greater estimated ability by raising the salary. A minister who improves relatively to his colleagues may expect prompt results in a better appointment. The results of deterioration will be equally prompt. The chief exceptions to this general freedom of competition of parishes for better ministers and ministers for better parishes which concern age are as follows:

(1) The bishops and district superintendents are supposed on the whole to favor the older men, delaying the rate of their decline at the expense of the rate of advance which young men would have if the choices of the parishes were always decisive.

(2) There have been provisions for allowances to ministers upon permanent retirement. These began as informal doles, made by a committee in each conference from funds collected annually and from the profits of the church publishing house. From 1896 to 1908, the rule of giving $10 annually for each year of active service was followed more or less generally. From 1908 on, the expectation has been that one-seventieth of the average salary, exclusive of parsonage, for each year of active service, should be paid annually, a widow receiving half

[1] The number of parishes and ministers is of course not absolutely the same at any given date. New churches are established and old churches are given up or combined; and deaths, retirements, permanent and temporary, assignments to non-pastoral work, and the like, are not balanced exactly by the admission of new men. There is commonly a slight excess of parishes over ministers, this excess being taken care of usually by appointing students in theological schools to certain parishes.

of this amount. It is not clear how often this expectation has been fulfilled. The individual records which we have studied are rather irregular. Payment has depended upon annual collections; the allowance has been considered to some extent a charity; ministers and their widows are still encouraged to waive it. We are consequently unable to judge what the actual results of pension expectations have been in the case of the individuals studied. It seems probable that they have had little effect beyond accentuating the natural tendency to retire rather than accept a post which is far inferior in salary and dignity to the posts held during one's prime.

(3) There are some cases where the individuals concerned have other sources of income and so are free to indulge predilections for special sorts of work, as in missions, or are so devoted to certain aspects of church work as to be careless of the matter of income for themselves and their families. However, the size of the salary is very closely correlated with the general dignity and importance of the post; and we believe that at least 95 per cent of the men in our records were impelled by personal ambition and family needs to desire as good salaries as they could obtain. So far as the relation of salary to age is concerned, the exceptional cases will cause little distortion of any general relation; for the man who has a private fortune will usually have it (or the expectation of it) for most of the years from 30 to 70, and the man impelled by special devotion will usually be impelled by it for most of these years.

We may then expect the salary records of ministers who live till 70, to show with approximate truth facts about the relation of earning power to age, the locus and extent of the "prime" of life, and the nature of the change from it toward the weakness of old age.

The cases selected for study were all the ministers who were, at the time of their death, members of the Maine, East Maine, New York, New York East, and New England Annual Conferences of the Methodist Episcopal Church, provided their death occurred between the years 1905 and 1924. (In the Maine and East Maine conferences, cases where death occurred between 1900 and 1905 were also included.)

There were utilized records of men who had spent their entire ministerial life within the bounds of the one conference and, as well, those who had been members of other conferences for a part of their career. The records of the following conferences in addition to the five used in entirety were thus also used in the study: New England Southern, Genesee, Oneida, Western New York, Central New York, Lexington, Ohio, East Ohio, Kansas, South Kansas, Northwest Kansas, Southwest Kansas, Rock River, Illinois, Indiana, Erie, Newark, New

Jersey, Montana, California, Puget Sound, Michigan, New Hampshire, Southern California, Cincinnati, Northern New York, Philadelphia, Delaware, West Virginia, N. W. Swedish, Alabama, South Carolina, Wyoming, Nebraska, Colorado, St. Louis German, Utah Mission, Iowa, N. W. Iowa, Missouri, St. Louis, Arkansas, Vermont.

The records were taken from the minutes of these conferences from 1860 to 1925, except that in the New York East Conference, no records of salary earlier than 1865 were available. Salaries earlier than 1860 are not on the same basis as later ones, since prior to that date each minister received a fixed allowance governed in amount by the size of his family, the amount of travel necessary to cover his area, and the like. The General Conference of 1860 passed legislation permitting discrimination in salary in recognition of unusual achievement and differences in capacity to pay on the part of the churches. Thereafter the records of the Annual Conferences contained more or less complete data with reference to salary, allowance for house rent, apportionments and collections for the salary of presiding elders (later called district superintendents) and bishops.

In some cases the data were complicated by the fact that there had been recorded in one sum the salary of the minister, the salaries of the presiding elder and of the bishop, and the amount of the presumptive salary allowed as house rent for the parsonage furnished to the minister by the church. In some cases the data were incomplete because of omissions in the tables of salaries. Sometimes ministers changed appointments during the year, the division of salary between the two incumbents of one charge being uncertain. When a minister transferred from one conference meeting in the spring to one held in the fall, or *vice versa*, it was sometimes impossible to find the salary for the year when the transfer occurred. Occasionally he appeared to be receiving a salary in both conferences, but usually he seemed to have none in either.

There is a further source of error. In every case the appointments to charges (parishes) for the succeeding year would be made at the session of the annual conference, and be listed by charges and by the names of the ministers. The salary paid for that year would appear in the minutes of the next year's conference, listed, in about one-third of the copies of the minutes examined, only by the charges and the salaries. In case of an interim appointment because of transfer to another conference, or breakdown in health necessitating a relief from duty, or a shift in appointments due to promotions to fill vacancies, the minister previously appointed would be credited with the salary paid whether he received it or not. Such errors would not be frequent.

however, since in case of such a transfer or of a shift in charges within the conference, the following year's minutes would record it. Search was made in such cases to learn at what time the transfer occurred and if possible what was the salary division between the two incumbents. In case of a disability extending into the following year, similar notice would appear in the lack of appointment for that year, and care was taken to learn the precise amount paid, by using the presiding elders' reports of the year's work on their districts and any other clues that the minutes might contain.

In different years and in different conferences, differing methods of reporting data were used. Sometimes the salary and parsonage rental allowance were reported together, sometimes separately. Sometimes both the salary promised and the salary paid were listed, sometimes only the salary promised and the deficiency in case of incomplete payment. In all but a very few cases it was possible to be certain as to the figure given, but in those few cases it was impossible to tell whether the figure in the salary column represented that amount inclusive or exclusive of the amount specified in the rent column.

The data otherwise are subject only to such inaccuracies due to typographical errors and to faulty reporting, as seem inevitable in any collection of statistics. It is to be noted that the tables published even in the sixties and seventies, show surprisingly few evident errors. It is probably true that greater accuracy is not to be found in any other data in the United States reaching back to 1860 and depending upon a large number of individuals for collection and preservation.

The record taken for each individual included name, date of birth, date of death, age at death, number of years of ministerial service, date of entering the ministry, and appointment and salary for each year from the year of entrance until no further salary was received because of death or permanent withdrawal from active service. If a minister lapsed from active service for a time, his record was resumed at the time of his re-entry and appointment to a charge.[1]

As a result we have several hundred records of the sort shown in the first four columns of Table I. From these there were selected all the individuals who lived to be at least 70, and whose lives from 30 or earlier until permanent retirement at 53 or later, were spent (except for temporary retirements because of illness or some work as chaplain, teacher, missionary, or the like) as ministers or district superintendents

[1] We are deeply indebted to the Reverend Oliver S. Baketel, D.D., Editor of the *General Minutes* and Editor of the *Methodist Year Book*, for permission to obtain data from the files in the library of the Methodist Book Concern, 150 Fifth Avenue, New York City.

in the Methodist Episcopal Church. Cases of permanent retirements before 53 were arbitrarily excluded because of the possibility that the men then undertook some other remunerative work.

TABLE I

SAMPLE RECORD OF FACTS AS TRANSCRIBED FROM THE MINUTES OF VARIOUS CONFERENCES OF THE METHODIST EPISCOPAL CHURCH (COLUMNS I TO IV), AND AS TRANSMUTED INTO PERCENTAGES OF THE MEDIAN SALARY AT AGE 40 TO 50 (COLUMN VI) *

I	II	III	IV	V	VI
Date	Age	Cash salary	Estimated value of rent of parsonage	Corrected salary in terms of 1898 to 1902 values	Corrected salary as a per cent of the man's median corrected salary at age 40 to age 50
1877.....	30	$1,300	$600	$1,365	$94
1878.....	31	1,500	600	1,665	114
1879.....	32	1,500	600	1,755	120
1880.....	33	1,500	600	1,860	127
1881.....	34	1,672	?	1,990	136
1882.....	35	1,750	?	2,030	139
1883.....	36	1,750	?	1,978	135
1884.....	37	none		none	?
1885.....	38	none		none	?
1886.....	39	none		none	?
1887.....	40	none		none	?
1888.....	41	none		none	?
1889.....	42	none		none	?
1890.....	43	1,520	315	1,535	105
1891.....	44	1,500	300	1,500	103
1892.....	45	1,500	300	1,485	102
1893.....	46	1,500	300	1,470	101
1894.....	47	1,500	300	1,445	100
1895.....	48	1,200	500	1,170	84
1896.....	49	1,400	500	1,372	94
1897.....	50	1,400	500	1,383	95
1898.....	51	1,600	500	1,600	110
1899.....	52	1,600	500	1,600	110
1900.....	53	1,500	500	1,515	103
1901.....	54	900	400	900	62
1902.....	55	1,000	250	970	66
1903.....	56	? (part of year)		low	low
1904.....	57	none		none	...
1905.....	58	none		none	...
1906.....	59	none		none	...
1907.....	60	none		none	...
1908.....	61	none		none	...
1909.....	62	600		538	37
1910.....	63	? (part of year)		low	low
1911.....	64	retired permanently		none	...
1912.....	65	retired permanently		none	...
1913.....	66	retired permanently		none	...
1914.....	67	retired permanently		none	...
1915.....	68	retired permanently		none	...
1916.....	69	retired permanently		none	...
1917.....	70	retired permanently		none	...

* This record is not typical, being chosen to illustrate the irregularities rather than the general facts.

The salary item used in each case was the cash salary received. Besides this cash salary, the minister usually receives also the use of an unfurnished house, the rental value of which would be about a fourth of the cash salary. It would perhaps have been better to have included the estimated rental value, but this introduces a subjective factor in

the form of the different rental values which might be set upon houses of equal actual value by the "official boards" of different parishes. Moreover, there are the cases of houses correctly valued at, say, $1,000 a year, which are of actually less real use to a minister receiving $2,000 than a $500 house would be. If he has to heat and keep clean a house with larger rooms than he wants, the $2,000 cash plus a $1,000 house may be actually less desirable than $2,000 cash plus a $500 house.

In cases where the records state only the total of the cash payment and the estimated rental value of the parsonage, we have deducted 20 per cent of the total to obtain the probable cash payment.

Two thousand dollars in 1924 obviously means a smaller real wage than $2,000 in 1894 or 1904. To make the salaries of one year comparable with those of any other, we might divide each by some index number expressing the cost of living for that year, if the average or median salary curve from 1860 to 1924 followed the cost of living curve at all closely. But it does not. The facts are shown in Chart I. There are very great departures in the way of lag and reduction. It seems much better, consequently, to divide each salary for each year by a number such that the average or median salary for any one year will be the same as that for any other. We have done something which is approximately this, but even better for our purpose. We have divided the salary items in our records by such numbers as make the salary curve of Chart I a straight line parallel to the base line. In other words, we make the median salary in the New England Conference (which is almost co-extensive with Massachusetts) approximately the same from 1860 to 1924; and also make the median salary in the New York Conference (which includes the southeastern part of New York, except parts of New York City and Long Island) approximately the same from 1860 to 1924.

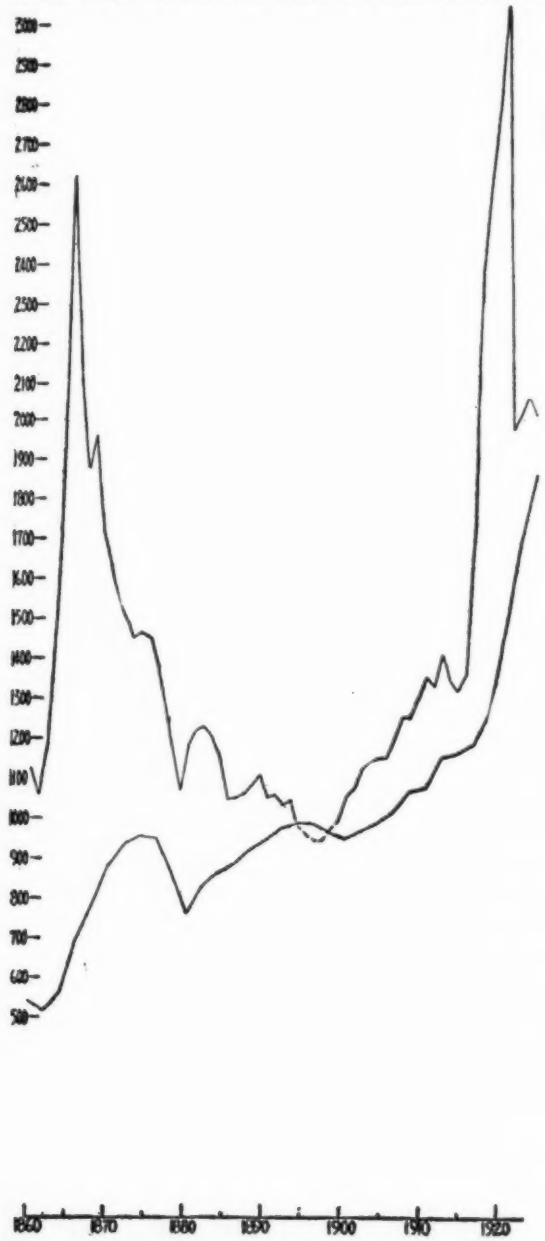
A still better procedure would be to use a separate divisor for each salary amount for each year, so that not only the medians, but also the variabilities, would be the same from 1860 to 1924. The computations necessary for this are, however, very tedious, and not necessary, since the 10-percentile salary curve, the 20-percentile salary curve, and so on to the 90-percentile salary curve, will all be straightened out by the divisors we have used to be nearly enough parallel to the median salary curve for our purpose.[1]

The divisors which we used are, for the years from 1860, in order, as shown in Table II.

[1] The facts in support of this statement will be presented in an article entitled "The Effect of Violent Price-fluctuations upon the Salaries of Clergymen" in a later number of this JOURNAL.

CHART I

THE CURVE OF THE COST OF COMMODITIES (ABOVE) AND THE CURVE OF MEDIAN SALARY EXCLUSIVE OF THE RENTAL VALUE OF THE PARSONAGE (BELOW)[1]



[1] The last zero of the scale is discarded to give the King index-number. The scale as it stands gives the salary in dollars.

TABLE II

DIVISORS FOR YEARS 1860 TO 1924 TO MAKE SALARIES COMPARABLE

1860	540	1882	830	1904	990
1861	530	1883	848	1905	1,005
1862	525	1884	865	1906	1,020
1863	540	1885	880	1907	1,035
1864	565	1886	895	1908	1,050
1865	620	1887	908	1909	1,070
1866	675	1888	925	1910	1,090
1867	735	1889	938	1911	1,110
1868	785	1890	950	1912	1,130
1869	835	1891	960	1913	1,148
1870	880	1892	970	1914	1,165
1871	908	1893	980	1915	1,182
1872	935	1894	990	1916	1,200
1873	945	1895	985	1917	1,240
1874	955	1896	980	1918	1,280
1875	950	1897	972	1919	1,370
1876	945	1898	965	1920	1,470
1877	915	1899	957	1921	1,585
1878	868	1900	950	1922	1,700
1879	820	1901	960	1923	1,785
1880	775	1902	970	1924	1,865
1881	805	1903	980		

We have then for each individual a series of amounts, one for each year, representing approximately each salary which he would have received if salaries in general had been from 1860 to 1924 at the average level of 1898 to 1902. For example, the record of individual G. L. C. shown in Columns I to IV of Table I assumes the form shown in the fifth column of Table I.

The next matter to be adjusted concerns the treatment of years for which records are lacking. In some such cases the man really earned nothing in the year in question. In others he probably earned even more than he would have earned ordinarily, by a special post in some special church enterprise, or in teaching, or as chaplain in the army. If the records show that he served some church as pastor, only the salary fact being lacking, we have estimated his salary as best we could from the records of what that church at about that time usually paid, and what that man before and after that year received, or have left the record as *w* meaning working, but for an unknown salary.

If the record states that the man received no appointment from the Methodist Church, we leave a dash on the entry and treat it in connection with others in a group by methods to be described later. In a few such cases the man may have earned something by some other occupation, but the great majority of them are temporary retirements because of illness, or permanent retirements in old age. If we have no information, as is sometimes the case in the early records, a question is entered.

We have then a record for each man for each year of the probable cash salary which he would have received if salaries in general had been always at the level of the average for 1898–1902 (corrected salary) or of the fact that he was working, the salary being unknown (*w*), or of

the fact that he was doing no work for payment for the church (–). Each entry of a man's record is then expressed as a per cent of the median of the corrected salaries received by him from age 40 to age 50 inclusive. The purpose of this last transmutation is to make the individuals more comparable in respect of the relation of salary to age.

The records so transmuted are grouped into four classes:

1 to 60	men whose	median	corrected 40–50	cash salary	is under $1,000
100 to 134	"	"	"	"	from $1,000 to $1,499
150 to 174	"	"	"	"	" $1,500 to $1,999
200 to 222	"	"	"	"	" $2,000 or over

Table III shows the records of cases 21 to 30 from the first group. The entries of such records will be called "percentile earnings."

Any one individual's record may be distorted here and there by errors due to the causes described earlier, and by eccentric cases of estimated rental value. For example, in certain city churches the estimated rental value may equal the entire cash salary, so that our deduction of 20 per cent may leave a cash salary far above what the man really received unless he was permitted to rent the house to somebody and live elsewhere himself. Some very low salaries may be due to the fact, suspected but not demonstrable by us, that the man really worked only a part of the year there and did earn more by working elsewhere. The general drift in the group as a whole may, however, be determined with sufficient exactitude, by using medians which will be little influenced by such errors.

In computing these medians our procedure has been as follows.

All records of *w* or ? are treated as if they represented a random sampling of the men who did receive salaries from parishes. That is, they are omitted from the computations. A temporary retirement when the man is known to receive nothing from the church is always rated as considerably below what the median percentile earnings of the group for that age would be if all the men worked at normal capacity. It is in fact probably usually near zero, representing illness or a vacation. That is, the median percentile earnings is computed for any age in any group by counting all the (–) entries as "low." This procedure is in error only by shifting the median percentile earnings a bit too low when a man retires temporarily to engage in some remunerative employment nearly or quite as profitable as his continuance in the ministry would have been.

A permanent retirement after the age of 53 when the man is known to receive nothing (save his pension allowance) from the church is rated as zero. This is not accurate, since some of these men doubtless earned something by preaching for parishes whose regular ministers

TABLE III

SALARIES OF TEN INDIVIDUALS FROM AGE 30 TO AGE 70, EACH BEING EXPRESSED AS A PER CENT OF THE MEDIAN OF THE SALARIES WHICH HE WOULD HAVE RECEIVED FROM AGE 40 TO AGE 50 IF SALARIES IN GENERAL HAD REMAINED AT THE LEVEL OF 1898 TO 1902

	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
21.....	*w*	*w*	*w*	*w*	*w*	*w*	*w*	44	41	115	82	71	69	102	84
22.....	72	83	65	100	112	136	116	108	101	96	96	95	120	72	100
23.....	81	124	120	137	123	109	106	97	93	88	*w*	*w*	*w*	109	109
24.....	63	*w*	*w*	*w*	*w*	103	106	98	130	123	103	100	100	123	122
25.....	?	?	?	97	108	97	50	94	116	113	–	–	–	*w*	*w*
26.....	218	203	116	141	116	116	108	102	104	*w*	*w*	111	110	110	100
27.....	64	*w*	*w*	*w*	87	79	70	75	71	69	80	80	73	101	121
28.....	62	88	84	91	83	121	*w*	*w*	*w*	104	85	90	89	85	93
29.....	33	72	69	71	69	63	57	56	68	68	77	82	92	98	99
30.....	107	116	108	100	93	88	111	90	87	78	86	90	91	93	99

	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
21.....	84	127	110	111	117	155	129	222	169	166	118	116	115	–	108
22.....	105	108	98	99	100	115	107	100	94	93	98	117	116	113	81
23.....	96	99	86	97	100	108	106	134	131	129	127	100	123	103	102
24.....	100	109	97	79	92	105	117	114	101	120	129	86	84	83	82
25.....	117	76	97	103	83	104	126	106	80	–	74	64	88	83	82
26.....	100	89	87	97	97	138	134	132	130	127	108	106	104	120	102
27.....	113	104	98	101	100	90	88	107	78	77	79	87	86	85	84
28.....	100	106	112	110	92	103	100	98	88	82	82	88	94	93	131
29.....	102	107	85	115	120	117	114	112	132	129	128	147	144	153	152
30.....	119	106	102	110	143	140	138	116	117	115	113	112	111	109	108

	60	61	62	63	64	65	66	67	68	69	70
21.....	71	104	103	–	–	–	–	–	–	–	–
22.....	110	91	–	–	–	–	–	–	–	–	–
23.....	101	113	116	101	115	99	83	–	–	–	–
24.....	81	80	80	59	59	79	80	81	*w*	*w*	*w*
25.....	81	–	–	–	–	–	–	–	–	–	–
26.....	101	100	92	93	104	108	102	–	–	–	–
27.....	80	–	–	–	–	–	–	–	–	–	–
28.....	137	129	120	137	128	99	96	59	59	59	59
29.....	150	148	98	89	89	120	121	101	102	101	–
30.....	92	92	93	94	92	92	96	95	–	–	–

were ill or on vacation, or by small farming or the like. Moreover, a few of them may have undertaken regular employment for wages outside the church. The influence of these errors upon the medians is nearly or quite zero until the number of "dash" records becomes greater than 50 per cent of the number of dash and salary records. Then the computation gives zero as the median man's percentile earnings, and this zero should be interpreted simply as "very low."

Table IV and Chart II show the percentile earnings of the median man in each of the four groups in relation to age. Table V and Chart III show the same facts put back in terms of dollars for men whose median cash corrected salaries from age 40 to age 50 were 800, 1,200, 1,700, and 2,500 respectively. The zero or "very low" medians for

TABLE IV

MEDIANS BY TWO-YEAR PERIODS OF SALARIES EXPRESSED AS IN TABLE III

The Entries at Age 70 are for a Single Year

Group	*N*	Median corrected salary, age 40–50	30 31	32 33	34 35	36 37	38 39	40 41	42 43	44 45	46 47	48 49	50 51
I	60	Under $1,000....	82.5	88	96	96	99.5	98	101	101	100	100	100
II	35	$1,000 to $1,499..	67.5	77.5	83	86	90	98.5	104	100	99	99.5	100
III	23	$1,500 to $1,999..	60	63.5	70.5	66	77	87.5	97	100.5	100	102	104
IV	23	$2,000 and over..	51	57	65	77	85	99.5	102	101	100	100	99

Group	*N*	Median corrected salary, age 40–50	52 53	54 55	56 57	58 59	60 61	62 63	64 65	66 67	68 69	70
I	60	Under $1,000	100	95.5	94	89	85	81	72.5	42	–	–
II	35	$1,000 to $1,499..	97.5	106	104	98.5	85.5	74	73	62.5	40	–
III	23	$1,500 to $1,999..	104	106	99	92	82	73	66	54.5	50	46
IV	23	$2,000 and over..	97.5	96	88	88	89.5	70	64	55.5	35	–

CHART II

PERCENTILE EARNINGS IN RELATION TO AGE

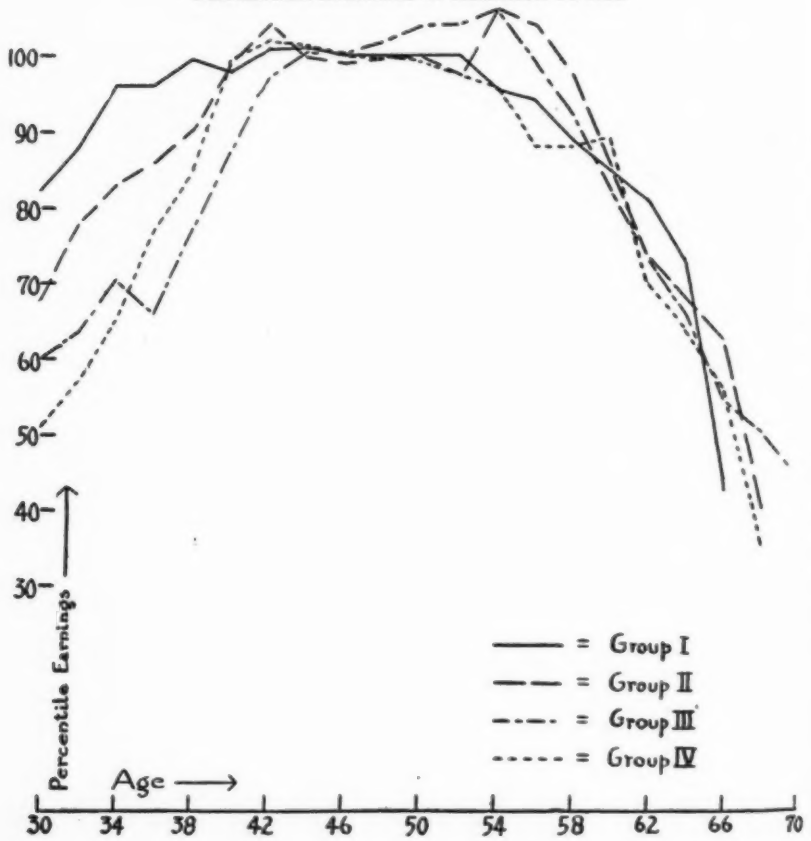


TABLE V

ESTIMATED MEDIAN SALARY IN RELATION TO AGE IN TERMS OF 1898–1902 VALUES FOR MEN WHOSE MEDIAN SALARY AT AGE 40–50 IN TERMS OF 1898–1902 VALUES IS $800, OR $1,200, OR $1,700, OR $2,500

Median Salary, Age 40–50	30–31	32–33	34–35	36–37	38–39	40–41	42–43	44–45	46–47	48–49	50–51
$800	660	704	768	768	796	784	808	808	800	800	800
$1,200	810	930	996	1,032	1,080	1,182	1,248	1,200	1,188	1,194	1,200
$1,700	1,020	1,080	1,199	1,122	1,309	1,488	1,649	1,709	1,700	1,734	1,768
$2,500	1,275	1,425	1,625	1,925	2,125	2,488	2,550	2,525	2,500	2,500	2,475

Median Salary, Age 40–50	52–53	54–55	56–57	58–59	60–61	62–63	64–65	66–67	68–69	70
$800	800	764	752	712	680	649	580	336	–	–
$1,200	1,170	1,272	1,248	1,182	1,026	1,008	996	750	480	–
$1,700	1,768	1,820	1,683	1,564	1,394	1,241	1,122	927	850	782
$2,500	2,438	2,400	2,200	2,200	2,238	1,750	1,600	1,388	875	–

age 70 in groups I, II, and IV, and for 68–69 in group II are not entered in the graphs.

The abler men rise from lower percentile earnings at 30, making much greater gains in dollars from 30 to 40. There is perhaps a slight tendency for the abler men to have their salary acme a little later than the less able, but this is so slight that it may be neglected. The lowest group retire permanently at younger ages than the others, presumably because the quarter or third of their 40–50 salary which might be offered them seems not worth its cost in loss of pride and labor, and because there are few parishes which would prefer to pay two or three hundred dollars (values of 1900) for their services rather than four or five hundred for those of a younger or abler man.

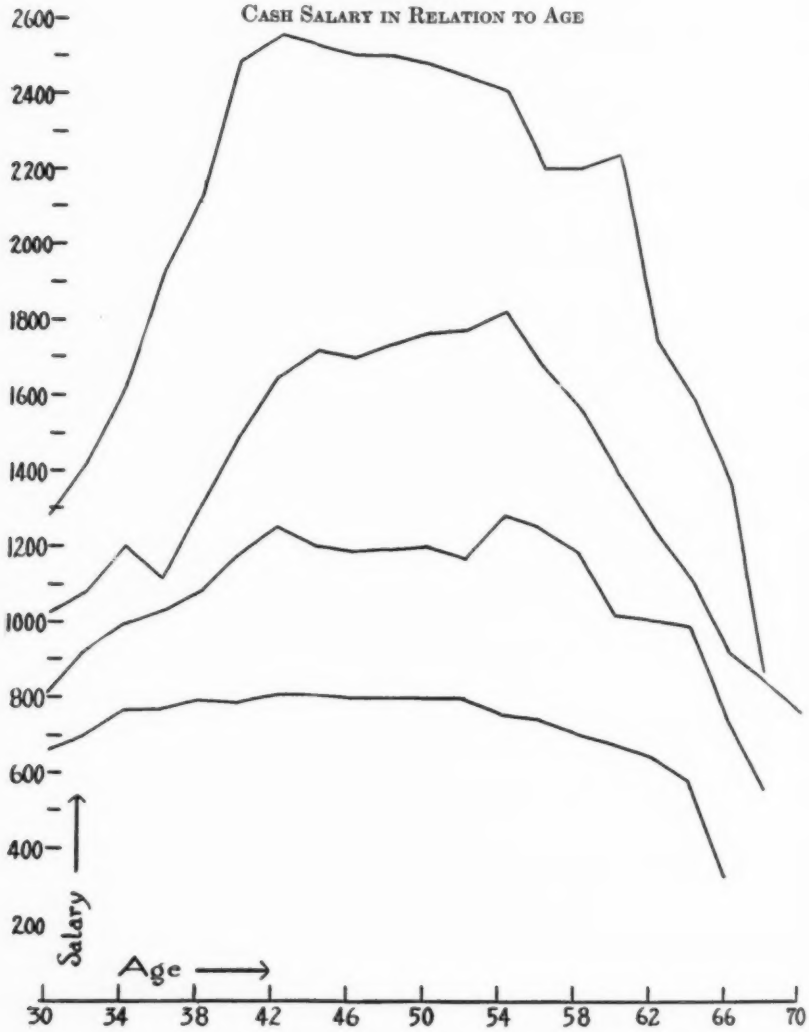
For all four groups the change in earnings from about 40 to about 56 is very, very slight. The prime of life is not a mountain, but a plateau. The drop thereafter increases with age. But it is noteworthy that with as free competition as prevails in respect of these salaries, the men in the three upper groups receive at about 66 as high salaries as they did at 31.

Individuals differ enormously from the general drift of the group. For example, there are entirely authentic records of men receiving more per year at 66 to 70 than they received from 40 to 50. The years which are a plateau for the median person may be a notable rise for one and a decline for another. Systems of promotion and retirement which depend in any rigid way on age are apparently destined to be misfits for very many.

The record for any one person for one year may, as has been stated,

CHART III

CASH SALARY IN RELATION TO AGE



be distorted considerably by errors, but if we use the median percentile earnings for each man for the 9 years 51 to 59 and for the 8 years 60 to 67, we shall be substantially free from these errors. Table VI presents these facts for the individuals of groups II and III, and also the records for ages 68, 69, and 70, which are of special interest in relation to systems of pensions. The entries for the median of the years 40 to 50 would of course all be 100.

Since there are no demonstrable differences in relation to ability in

TABLE VI

INDIVIDUAL DIFFERENCES IN THE AGE SALARY RELATION AFTER 50

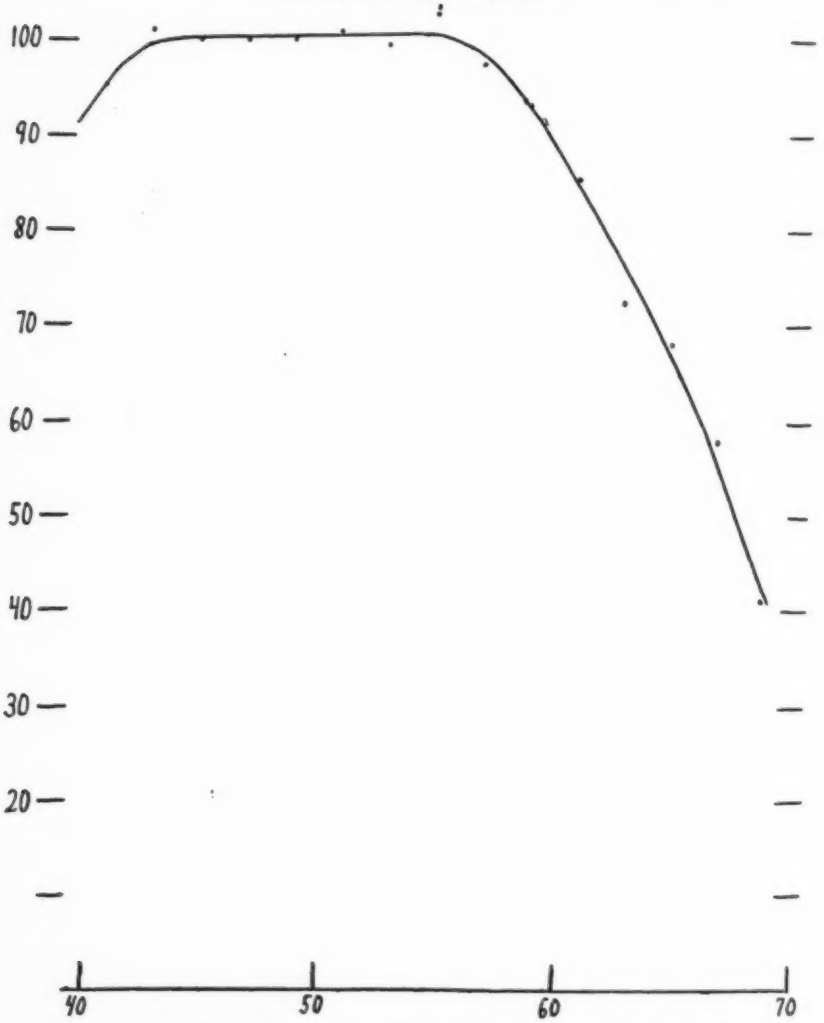
Individual	Median 51–59	Median 60–67	68	69	70
Group II					
101	89	66.5	50	46	..
102	96.5	<40	40	40	..
103	115	91	78	60	..
104	108	79.5	79	80	80
105	106	94.5	56	52	85
106	..	..	..	..	..
107	185	111	140	140	143
108	80	78	79	40	59
109	111	?	*w*	*w*	*w*
110	67	65	..	..	..
111	142	138	104	70	104
112	117	74.5	63	62	86
113	103	95.5	97	48	83
114	107	72.5	..	..	..
115	..	..	..	..	..
116	156	112.5	*w*	*w*	*w*
117	112	87.5	67	66	66
118	90	129	118	69	69
119	128	88.5	?	43	43
120	81	66	..	..	..
121	116	100	*w*	*w*	*w*
122	96	101	117	112	33
123	83	75	78	78	77
125*	105	..	..	..	..
126	72	..	..	..	..
127	150	..	..	..	..
128	154	92	68	..	..
129	127	118	74	73	54
130	77	52	51	..	..
131	109	<52	*w*	..	..
132	65	..	..	..	..
133	121	76	..	..	..
134	62	..	..	..	..
135	83	..	..	..	..
136	..	<62	60	..	..
Group III					
151	..	51	51	52	52
152	154	135	173	182	160
153	102	98.5	104	106	95
154	51	37	50	..	..
155	106	..	..	..	..
156	70	67.5	48	48	48
157	102	63.5	*w*	..	..
158	105	76	..	..	..
159	116	110	58	57	56
161*	100	79	69	68	54
162	102	78	51	51	55
163	*w*	39	33	35	34
164	109	121.5	..	..	..
165	92	62	52	..	..
166	108	55.5	..	..	..
167	110.5	66.5	42	41	46
168	129	130	170	67	88
169	90	74	52	51	50
170	64	34.5	59	60	44
171	119	79	66	49	64
172	93	71	..	..	..
173	61	52	..	..	..
174	108	84.5	77	80	60

*There are no individuals 124 and 160.

the three upper groups after age 40,[1] we may combine the records and compute more reliable measures[2] for the Methodist minister receiving at his prime a cash salary of $1,000 or more, at values of 1898–1902. The results are presented in Table VII and Chart IV, which, subject

CHART IV

THE GENERAL RELATION OF AGE TO THE ESTIMATED WORTH OF THE WORK OF METHODIST MINISTERS RECEIVING SALARIES OF $1,000 OR OVER, EXCLUSIVE OF PARSONAGE



[1] Such differences as appear are well within the probable errors of the medians themselves.

[2] We have used weighted averages of the three medians, giving weights of 4, 3, and 3 to II, III, and IV respectively.

to improvement by more extensive investigations, are a reasonable estimate of the general relation of age to the estimated worth of the services of a man doing the sort of work done by a Methodist minister. The relation in any individual person may, as has been shown, vary widely from this.

TABLE VII

THE RELATION OF SALARY TO AGE FOR MINISTERS WITH SALARIES (EXCLUSIVE OF PARSONAGE) FOR $1,000 AND OVER AT AGES 40 TO 50

Salaries are expressed as per cents of the median salary at 40 to 50

Age	Percentile Earnings
40 and 41	96
42 and 43	101
44 and 45	100
46 and 47	100
48 and 49	100
50 and 51	101
52 and 53	99
54 and 55	103
56 and 57	98
58 and 59	94
60 and 61	86
62 and 63	73
64 and 65	68
66 and 67	58
68 and 69	42

THE DETERMINATION OF CURVILINEAR REGRESSION "SURFACES" IN THE PRESENCE OF OTHER VARIABLES

By Mordecai Ezekiel, *Bureau of Agricultural Economics, United States Department of Agriculture*

INTRODUCTION

By the use of multiple correlation, one effect may be explained as due to a number of different causes, splitting up the variation in the dependent variable and accounting for one part of it by variation in one variable, another part by variation in another, and so on. Even where the variation is stated in terms of relative rather than absolute relations,[1] it still holds true that multiple correlation provides a method whereby the variation of the dependent variable can be distributed in such a way that parts of the variation can be separately imputed to each of the independent variables, and part to residual causes.

In actual fact, however, many relations are of such type that they cannot be treated as due to the separate action of even a number of variables, but must be considered as due to the joint action of all of them. Examples of such a nature may be readily drawn from biology. The rapidity of the growth of a seedling, for example, is dependent upon at least four primary factors: Weight of the seed, moisture, heat, and light. Now with one combination of seed, heat, and light, certain differences in moisture will result in certain differences in the rate of growth; with a different higher temperature, but the same seed and light, the same changes in moisture will result in certain other changes in growth; while with a temperature lower than the original, the same changes in moisture will result in still different changes in growth. That is to say, a certain change in moisture does not result in a certain change in growth, regardless of the combination at which the other factors are held constant by experimental means; but the very relation between changes in moisture and changes in growth itself varies at different temperatures.

This same result will hold true for other combinations of the factors. The effect of changes in light on changes in growth will be different for plants receiving much water than for plants receiving little. Even the effect of differences in the weight of the seed upon the growth of

[1] Mordecai Ezekiel, "The Assumptions Implied in the Multiple Regression Equation," this Journal, September 1925.

the plant will differ as between experiments with the other conditions held constant at one set of relations, or held constant at another set of relations.

Stated in other terminology, holding three factors constant by experimental means and varying the fourth gives a different regression for one combination of the constant variables than it gives for another combination of the constant variables. The net change in the dependent factor for changes in the fourth independent factor varies according to the particular combination of values at which the three other independent factors are held constant. For this reason, relations of this type cannot be adequately expressed by multiple regression equations even of the most general type, such as

$$X_1 = f_2(X_2) + f_3(X_3) + f_4(X_4) + f_5(X_5) + A$$

or

$$X_1 = K[\Theta_2(X_2)][\Theta_3(X_3)][\Theta_4(X_4)][\Theta_5(X_5)].$$

But must be expressed by some such relation as

$$X_1 = f(X_2,\ X_3,\ X_4,\ X_5).$$

Stated in words, this means that given values of X_1 would have to be explained as due to the joint effect of specific values of the other variables, there being possible as many individual estimates of the value of X_1 as there are different combinations of the values of the independent variables.

However, the attempt to determine average values of X_1 for every possible combination of the independent variables would involve almost insuperable difficulties in problems where many variables are considered, unless practically unlimited observations are available. Even then, the observed relation for one set of values would not necessarily bear any consistent relation to a combination but one change removed.

At this point we seem to have lost entirely the element of generalization which is the fundamental concept in correlation, and are merely treating minute classification of data without making any attempt to block out the general sweep of the relationships. With sufficient data, however, we could make averages of X_1 for every change in X_2 for one given combination of the other factors and smooth the line of these averages to such curve as seemed justified. This would then give a statement of the changes in X_1 with changes in X_2, for *given specific* values of X_3, X_4, and X_5. By repeating the process with various other combinations of X_3, X_4, and X_5, it would be possible to get a series of curves expressing the relation between X_1 and X_2, one curve for each specific combination of X_2, X_3, and X_4. But while theoretically feasi-

ble, for practical work the limits of material and time compel us to employ some other approach to the problem.

The following discussion starts with the consideration of a method of handling the simplest case, that where the dependent variable is considered as the joint function of two independent variables, and then develops methods by which such relationships may be measured even where the effects of other variables or similar groups of variables would have to be eliminated simultaneously.

ONE VARIABLE AS THE JOINT FUNCTION OF TWO OTHERS

One variable may be graphically expressed as a function of two others acting jointly by using a "solid" diagram on isometrically ruled paper. Such a relation is shown in Figure 1.

Here X_1, the dependent variable, is represented by the vertical scale; the independent variable X_2 by the horizontal scale A to D, and the independent variable X_3 by the horizontal scale A to C. The regression of X_1 on X_2 and X_3, $X_1 = f(X_2, X_3)$ is represented by the surface $A'\ C'\ F'\ D'$.

It is evident that an attempt to measure the relation by the use of the formula

$$X_1 = f_2(X_2) + f_3(X_3), \text{ or even the formula}$$
$$X_1 = [\Theta_2(X_2)][\Theta_3(X_3)]$$

could not be successful, since in the section $A\ D$ (X_3 constant), X_1 decreases as X_2 increases; in the section $B\ E$ (X_3 constant at a different value), X_1 remains constant as X_2 increases; and in the section $C\ F$ (X_3 constant at a third value), X_1 increases as X_2 increases. Stated mathematically, the partial derivative of X_1 with respect to X_2 is different for each different value of X_3. Similarly, the partial derivative of X_1 with respect to X_3 is different for each different value of X_2.

The best approximation to this surface that could be obtained by measuring separately the net regressions of X_1 on X_2 and on X_3 would show for the net regression of X_1 on X_2 a curve which approximately represented an average of the slopes of curves such as $A'\ D'$, $B'\ E'$, $C'\ F'$, curves representing the intersections of the regression surface with planes parallel to the X_1 and X_2 axes. The exact shape of this regression curve would depend upon the weighting of different parts of the surface according to the number of observations occurring at different combinations of the two independent variables.

Ordinarily, the greater part of the observations would be found near the averages. In this illustration the negative relation of X_1 to X_2 for values of X_3 less than B might be approximately offset by the positive

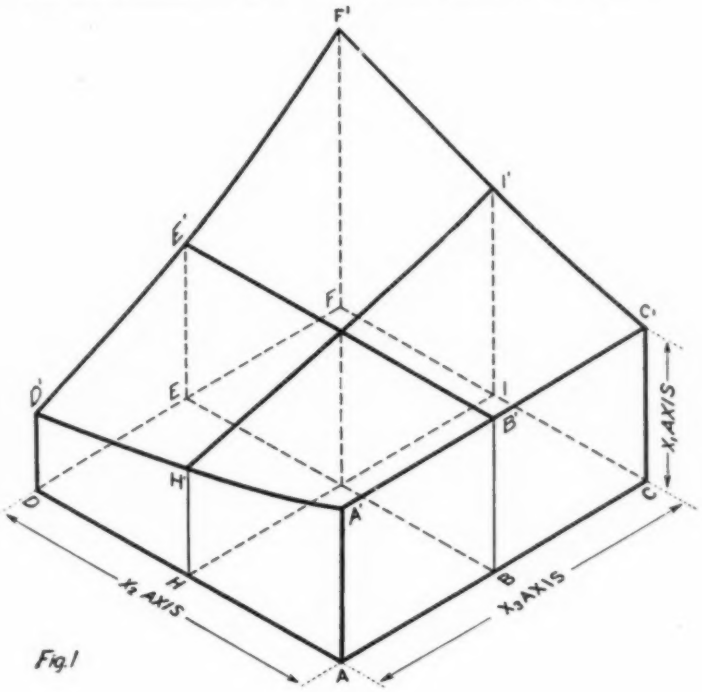

F'
I'
E'
F
C'
D'
E
I
B'
H'
A'
D
C
H
B
A
X_1 AXIS
X_2 AXIS
X_3 AXIS

Fig. 1

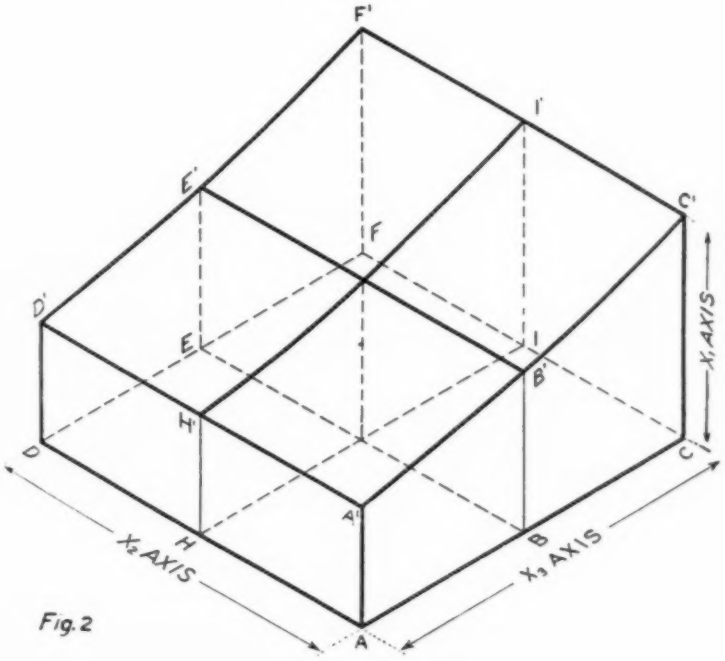

F'
I'
E'
C'
F
D'
E
I
B'
H'
A'
D
C
H
B
A
X_1 AXIS
X_2 AXIS
X_3 AXIS

Fig. 2

relation for values of X_3 greater than B, resulting in the net regression showing that there was no significant relation between X_1 and X_2. Similarly the curve for the net regression of X_1 on X_3 would be an approximate average of the different slopes such as $A'C'$, $D'F'$, etc., obtained by planes parallel to the X_1X_3 axes intersecting the surface.

Given data really having the regression surface shown in Figure 1, curvilinear correlation based on the assumption of a relation such as

$$X_1 = f_2(X_2) + f_3(X_3)$$

would give a net regression surface similar to the one shown in Figure 2. Here the partial derivative of X_1 with respect to X_2 is the same, regardless of the value of X_3, and the partial derivative of X_1 with respect to X_3 is the same, regardless of the value of X_2. That is, $(D'D)-(A'A)=(E'E)-(B'B)=(F'F)-(C'C)$; and similarly $(F'F)-(D'D)=(I'I)-(H'H)=(C'C)-(A'A)$. It is evident that this is but a very inadequate description of the relation shown in Figure 1, where $(C'C)-(A'A)<(F'F)-(D'D)$, and where $(D'D)<(A'A)$, $(E'E)=(B'B)$, and $(F'F)>(C'C)$.

It is probable that analytical expressions could be written for the relation shown in Figure 1 and constants determined by methods analogous to the method of least squares, so as to measure the relation from a set of observations more adequately than they can be measured by the net regression equation graphed in Figure 2. Such a solution would involve a very large amount of work and unless there was some definite theoretical basis for the particular analytical function used, might not give a much better fit than the ordinary regression equation. To meet this difficulty a method has been devised to determine the surface quite independently of analytical means or previous assumptions, obtaining the surface directly from the observations without the use of mathematical assumptions.

Where only two independent variables are concerned, as in Figure 1, the process is essentially two-way smoothing. The observations are grouped with regard to both X_2 and X_3, and the mean of X_1 taken for each group. The fineness of the grouping is of course dependent on the number of observations. Taking each X_2 array in turn, the X_1 means for successive X_3 groups in that array are plotted, and connected by continuous lines. Curves are then drawn free-hand through the broken lines thus formed. The ordinates of the curves are then read for selected X_3 values, and graphs made to show the smoothed value of X_1 with different values of X_2 for certain constant values of X_3. These values are then smoothed free-hand, new values read off, and the process of smoothing continued until a sufficiently regular surface

is obtained. (That is, the individual group-averages are smoothed by free-hand first in the plane $X_1\ X_3$, then these values are smoothed at right angles, in the plane $X_1\ X_2$, these again in the previous plane $X_1\ X_3$, and so on until a smooth surface is obtained.) Once the final regression surface is obtained, the correlation of this surface with the original observations may be obtained by either of two ways:

(1) Graph the surface on isometric paper, and scale off the regression-estimate of X_1 for each individual $X_2\ X_3$ combination observed; or (2) use the final smoothed X_1 value for each $X_2\ X_3$ group as the estimated value of X_1 for any $X_2\ X_3$ combination falling within the limits of that group. The first method is the more accurate, but if small enough groups are used the second method will give nearly as good results.

Having estimated the values of X_1 for each observation by either of these methods, all that remains to obtain the index of the multiple correlation of X_1 with X_2 and X_3, $P_{1\cdot 23}$, is to correlate these computed values with the original X_1 values. The same result may be secured more readily by simply computing the standard deviation of the original X_1 series and of the difference between the predicted or X'_1 series and the original series. Then letting $e' = X_1 - X'_1$, the index of multiple correlation is given by the formula[1]

$$P^2_{1\cdot 23} = 1 - \frac{\sigma^2_{e'}}{\sigma^2_{x_1}}.$$

Where only three variables are concerned the problem is hence simply one of reducing the irregular surface given by the group-averages to a smoothed surface, and then determining what part of the original variation in the dependent variable is accounted for by the joint function of the two independent variables represented by this surface.

Where additional variables are concerned, however, the problem is more complex. An example of such a problem would be one where the relations could be adequately expressed by a regression equation of the form $X_1 = f(X_2, X_3) + f(X_4) + f(X_5)$. (1)

Here the problem is not merely to determine the joint function represented in the regression of X_1 upon X_2 and X_3, but to determine this while eliminating or holding constant the effects of X_4 and X_5.

This may be approached by first considering variables X_2 and X_3 as separate independents, and solving by the method of multiple curvilinear correlation[2] to determine the curves for the relationship

[1] Frederick C. Mills, "The Measurement of Correlation and the Problem of Estimation," this JOURNAL, Vol. XIX, No. 147, p. 273. 1924.

Mordecai Ezekiel, "A Method of Handling Curvilinear Correlation for any Number of Variables," this JOURNAL, Vol. XIX, No. 148, p. 444. 1924.

[2] Mordecai Ezekiel, *loc. cit.*

$$X_1 = f'(X_2) + f'(X_3) + f'(X_4) + f'(X_5). \tag{2}$$

Having found this relationship, the effect of X_4 and X_5 on the dependent can be eliminated by computing a residual for each observation by the formula

$$Z' = X_1 - f'(X_4) - f'(X_5). \tag{3}$$

These residuals, Z', constitute the dependent variables with the net effect of X_4 and X_5 eliminated, to the extent to which their effect can be measured while using separate functions for the relation of X_1 to X_2 and X_3.

Then group these residuals according to the accompanying X_2 values, subdivide each of these groups according to the X_3 values, and average the values of Z' for each sub-group. The means of Z' for each of these sub-groups constitute a series of values giving an approximation to the net relation of X_1 to X_2 and X_3 together. These averages should then be "smoothed" in both directions by the method already outlined for two variable relations, and the smoothed surface used as the first approximation to the surface for the function of X_2 and X_3, $f'(X_2,X_3)$. Then from the new regression equation

$$X_1' = f'(X_2,X_3) + f'(X_4) + f'(X_5) \tag{4}$$

new residuals can be computed by the equation

$$e'' = X_1 - X_1' \tag{5}$$

and these residuals grouped in turn according to the accompanying values of X_4 and X_5 to make the next correction in the values of the f' curves. These new curves can then be used in a new equation

$$e''' = X_1 - f'(X_2,X_3) + f''(X_4) + f''(X_5). \tag{6}$$

These new residuals can then be grouped in turn according to the values of X_4 and X_5 to make further corrections in the f'' curves, and grouped and sub-grouped according to the values of X_2 and X_3 to make the next approximation correction for the surface $f'(X_2,X_3)$. A new residual can then be computed by the equation

$$e'''' = X_1 - f''(X_2,X_3) + f'''(X_4) + f'''(X_5) \tag{7}$$

and the series of approximations repeated as many times as seems desirable. A final check on the accuracy of the fit may be made in the same manner as suggested for simple curves by substituting the values read from the surface and curves for the original values of the independent variables, and correlating these values with the dependent.

That is, for given values of X_2 and X_3, substitute values, Y_2, determined so that $Y_2=f''(X_2,X_3)$
and for given values of X_4 and X_5 substitute, respectively, Y_4 and Y_5, where $Y_4=f''(X_4)$ and $Y_5=f''(X_5)$.

Then by multiple correlation determine the constants for the regression equation

$$X_1=a+b_2Y_2+b_4Y_4+b_5Y_5. \tag{8}$$

If the previous work has been well done, the constants b_2, b_4, and b_5 will be close to unity. Multiplying the values for the (f'') surface and curves by these constants gives the final values for the functions of equation (1). The coefficient of multiple correlation, $R_{x_1 \cdot y_2 y_4 y_5}$, is the index of multiple correlation of X_1 with X_2, X_3, X_4, and X_5, according to the relations shown in equation (1).

By a similar process of successive approximations the method could be applied to determine the regression surfaces where there were several involved, as for the relation expressed in the regression equation $X_1=f_2(X_2,X_3)+f_4(X_4,X_5)+f_6(X_6,X_7)$.

It would likewise be possible to use it to obtain regressions for relations such as

$$X_1=f_2(X_2,X_3,X_4)+f_5(X_5,X_6).$$

However, where the dependent was a function of three or more variables, the regression could not be shown graphically, since four dimensions would be needed [note the three dimensions in Figure 1 for the relation $X_1=f(X_2,X_3)$] but would have to be expressed merely as a series of tabular values, as suggested for the general case discussed in the introduction. Such three-variable functional relations could be determined only when a very large number of observations were available, as even if only five classes were used for each variable 125 sub-groups would have to be used in classing the residuals according to the three variables.

In a sense this detailed sub-grouping is a departure from the usual correlation point-of-view, in that it attempts to measure the *specific* effects of *specific* groupings of variables, instead of merely the average effect of each variable regardless of the other variables with which it may be combined. Where sufficient observations can be obtained, this determination of the specific relationship may prove to be the most important desideratum.

Obviously net regression surfaces could be obtained only if enough observations were available, and the accuracy of the different portions of the surface would depend upon the number of observations falling

within that portion. The remarks on the probable error of the curves for curvilinear correlation[1] would apply with equal force here. In considering the average residual for given subdivisions it is possible that the usual formula for the standard error of an average, $e_m = \frac{\sigma}{\sqrt{n}}$, might be applicable, considering the individual residuals as ordinary observations. Only when there was a fair degree of agreement among the residuals in a certain "cell," or only when there was a very large number of observations, would the use of this formula indicate that much significance could be attached to the average residual of the cases falling in that "cell."

ARITHMETIC EXAMPLE

The actual procedure in obtaining a net regression surface is so similar to that of obtaining net regression curves by the successive approximation method that it is not worth taking space to carry through an example in detail. The results from a problem worked out by this method will be presented, however, as an example of its possibilities.

In a study of milk production under farm conditions it was desired to find the relation between the feed the cows received and the quantity of milk they produced. Besides the feed record, figures were available as to the proportion of butter-fat in the milk, the proportion of the cows in each herd freshening in the fall (and hence milking during the winter season), and the value of the cows (as a measure of cow-quality).

The factors other than feed could not be considered as adding directly to the milk produced, but rather as modifying the production from a given quantity of feed, so a relative (logarithmic) regression equation was used.[2]

Letting X_1 = Annual milk production per cow
X_2 = Therms of net energy fed annually per cow, exclusive of pasture
X_3 = Proportion of net energy in the form of proteins
X_4 = Proportion of herd fall-freshened
X_5 = Butter-fat content of the milk
X_6 = Average value per cow[3]

and determining the curves for the regression equation

$$\log X_1 = f(\log X_2) + f(\log X_3) + f(\log X_4) + f(\log X_5) + f(\log X_6)$$

[1] Mordecai Ezekiel, *loc. cit.*, p. 449.

[2] Mordecai Ezekiel, "The Assumptions Implied in the Multiple Regression Equation," this JOURNAL, September 1925.

[3] The data consisted of averages from each of 354 herds, including about 5,000 cows in all. The methods of collecting the data are given in U. S. Department of Agriculture Bulletin 1400, *Factors Affecting Farmers' Earnings in Southeastern Pennsylvania*, by the present author.

a multiple correlation-index of .59 was obtained. On the assumption that the milk production was a function of specific combinations of net energy and protein, a regression equation of the form

$$\log X_1 = f(\log X_2, \log X_3) + f(\log X_4) + f(\log X_5) + f(\log X_6)$$

was tried. This raised the multiple correlation-index to $P = .61$.

While this total correlation was not sufficient to draw any very definite conclusions as to the exact effect of the various factors on the

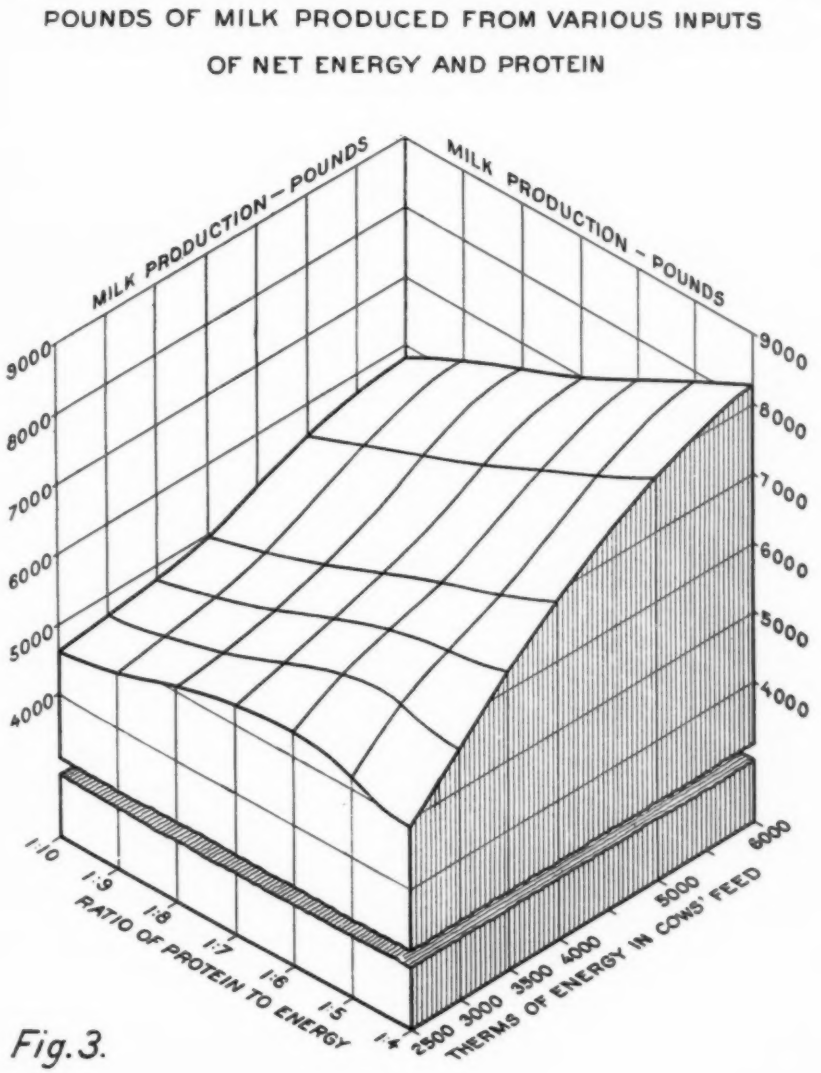


Fig. 3.

quantity of milk produced, the shape of the regression surface showing the net relation of specific combinations of energy and protein to milk production was of considerable interest. (Figure 3.) This surface indicated that the proportion of protein required in the feed to result in maximum production varied with the total input of energy, and was not a constant. While not conclusive, this introduced a specific problem which should be studied further both by controlled experiments and statistical investigation with more exact data.[1] But little examination of the figure is needed to show how poorly the relations would have been represented by two curves, one showing the average net relation of the quantity of energy fed to milk produced, and the other showing the average net relation of the proportion of protein in the feed to milk produced.

While the introduction of this further complexity in correlation analysis may be worth while only in exceptional problems, it goes one step further in making our analytical tool adequate to measure the complex relationships characteristic of the physical, economic, or psychical world.

[1] Since the data used in this study were obtained by the farm survey method, the figures were very rough in many cases. The correlation of .61 is fairly satisfactory in view of this inexactness of the original data.

THE TREND-SEASONAL NORMAL IN TIME SERIES

By C. H. Whelden, Jr., *Yale University*

A considerable literature has developed of late on the methods of measuring secular trend and seasonal variation in chronological series. It is now possible to make such measurements with as much precision as the character of the data will justify in most cases. The process of eliminating the influence of secular trend and seasonal variation from the raw data has not been so carefully examined. It is with this latter part of the problem of treating series so as to secure an indication of the effect of the so-called business cycle that the present discussion is concerned, coming from a conviction on the part of the writer that the method of eliminating trend and seasonal now commonly employed is in theory illogical, and in practical use a possible source of error.

The method of elimination which is almost universally employed at present, the method developed by Professor Persons in his pioneering study in the field of the analysis of time series, consists fundamentally, although with some variations of form, of determining an index of seasonal variation in terms of average annual activity and subtracting this index from the series of percentages obtained by dividing each original item by the corresponding ordinate of the line of secular trend. The process of the subtraction of the seasonal percentages from the percentages of trend is, as Professor Persons indicates, an arithmetical short-cut for the more exacting process of the division of one percentage by the other, the result being the same in this latter form as if the ordinate of trend were multiplied by the seasonal percentage and the original item of the series were then expressed as a percentage of this new trend-seasonal ordinate.

The criticism of this method rests upon the fact that the two sets of measurements which are combined are incommensurable, one being on the base of average annual activity, the other on the base of particular activities in particular sub-annual periods. The exact statement of the criticism must be reserved until after a brief examination of the character of an index of seasonal variation and of the meaning of normal activity in relation to the movements of the so-called cycle.

In a theoretically ideal world of activity, perfectly mobile in every part, perfectly foreseen in every aspect, devoid of unusual sporadic events, unaffected by changing seasons, shifting psychology and

tastes, or dancing dollars, the only variation in activity would be the regular movement arising from the organic evolution of human society. It is to this movement fundamentally we refer when we speak of secular trend, although in any particular case, of course, other elements of regular variation may be included in the particular secular trend. Whether the secular trend is fundamental or partially derived in the above sense, it is always a movement which may be designated as normal for any given period when the focus of attention is on movements which work themselves out in relatively brief portions of that period. It is, of course, a moving normal. The normal position for any one month or other sub-annual period differs from the average normal position for the year, except where the moving normal or line of trend is horizontal throughout. Furthermore, any two months equally distant from the middle of a given year, or the same months in any two distinct years will not have trend-normals which differ proportionally from the respective average normals for the years, unless the trend-normal is changing according to a law of simple proportions, that is logarithmically.

If we alter the conditions of the theoretical world of activity by admitting the influence of the round of the seasons, we shall have a regularly recurring seasonal variation superimposed on the underlying normal variation which we call trend. The only effect of this seasonal movement will be to redistribute throughout each year what would have been the amount of activity for that year if seasonal variation had not existed and the only variation had been that of trend. The total amount, and the average amount on any sub-annual base, of activity in any year will not be changed from the normal position by seasonal variation. The only change will be in the amount of activity found in any particular month, or other sub-annual period. Difference of activity from month to month will be a function of the movement not only of secular trend, but also of seasonal variation.

As the general amount of activity varies from year to year, the index of seasonal variation is necessarily in terms of percentages, and is most useful in terms of percentages of the average activity for any year, that is the average activity in any one sub-annual period. It is immaterial to the present discussion whether the series of percentages which constitutes the seasonal index remains constant from year to year, changes gradually, or shifts abruptly. The essential element is that in any year the seasonal index gives simply a seasonal distribution throughout the year of the average activity for that year. Seasonal variation is not properly considered as affecting the absolute amount of activity but only the relative amounts of the given ac-

tivity which occur in particular periods of the year. If the amount of activity in any year is taken as a standard of reference, the effect of seasonal variation cancels itself out in the course of the year, leaving the amount of activity in that year exactly what it would have been if seasonal variation had not existed. This point has been especially emphasized because on this point the present method of eliminating seasonal variation introduces error.

If now we abandon the theoretical world of activity and turn attention to the world as we know it with the desire of measuring the alteration in activity from that composite group of influences which we lump together for designation in the phrase "business cycle," the base from which the alteration must be measured, the normal to which the cyclical fluctuation is referred, is the composite trend-seasonal movement which would have existed if it had not been altered by the cyclical influence. The measure of trend and the measure of seasonal must be combined into a single series of base figures, and must be so combined that the sum of these base figures remains exactly the same as it would have been if there had been no seasonal variation acting in conjunction with the movement of trend.

The accepted method of eliminating from the raw data the effect of secular trend and seasonal variation and so leaving a measure of the cyclical and irregular influences is illogical and inaccurate, because in securing its trend-seasonal normal it so combines the elements of trend and seasonal that the trend-normal for any particular year is not merely redistributed over the seasonal periods of the year but is also altered, in varying degrees, in its absolute level for the year as a whole. The essential error comes from combining a seasonal index which is expressed in terms of sub-annual percentages of the average activity for the year with sub-annual ordinates of trend each one of which is different from the average of the trend activity for the year and differs ordinarily from that average in an irregular percentage relationship. The error is more obvious if it is remembered that the average annual activity on which the seasonal percentages are based, from the point of view of deriving a normal position from which to measure cyclical deviations, is simply the average of the trend activity for the year, that the seasonal figures are percentages of the average trend-normal for the year and not percentages of the trend-normals for the particular portions of the year. The following arithmetical examples will illustrate the error involved in the usual method and the type of procedure which logic demands.

Example A assumes for two years quarterly ordinates of a trend increasing arithmetically, assumes items of data which give the per

cent relatives to trend as 120 uniformly, and assumes a seasonal index, the same for both years, expressed in the customary form of percentages of average annual activity. The increment of trend and the extent of seasonal variation have been made unusually large in order to throw into relief the error involved in the accepted method of correcting the data for trend and seasonal. In order to allow the old method the full measure of its possible accuracy, the per cent relatives to trend are corrected for seasonal variation by dividing each by the corresponding seasonal figure rather than by the arithmetical short-cut of subtraction. The resulting figures (Column f) are supposed to represent the percentage relationship of the original items to the normal amount of activity for the respective quarters.

EXAMPLE A

Time	Item	Ordinates of Trend	Per cent Relative to Trend	Seasonal Index	Per cent of Normal	Indicated Normal
(a)	(b)	(c)	(d)	(e)	(f)	(g)
Year I						
1st quarter	12	10	120	120	100	12
2nd "	24	20	120	90	133	18
3rd "	36	30	120	110	109	33
4th "	48	40	120	80	150	32
		100				95
Year II						
1st quarter	60	50	120	120	100	60
2nd "	72	60	120	90	133	54
3rd "	84	70	120	110	109	77
4th "	96	80	120	80	150	64
		260				255

These percentages of normal (P), when taken with the corresponding original items (I), permit us to find the figures which represent normal (N) for each quarter by substituting in the equation, $I = PN$, or we can, under the method of elimination here used, obtain the same result by multiplying the ordinate of trend for each quarter by the corresponding percentage of seasonal variation. The results are given in Column g. They are supposed to be the figures which represent normal for this variable, that is the values which the variable would assume if there were no cyclical nor irregular fluctuations but only the tendency of trend overlaid by a seasonal variation. The ordinates of trend are the figures which would have represented normal if seasonal variation had also been absent. Since only the sub-annual distribution of the normal and not its total or average level for the year can, a priori, be affected by the presence of a seasonal variation, the average or total for each year of the ordinates of trend (Column c)

should be the same as the average or total for each year of the indicated normals (Column g). Obviously they are not: in Year I the indicated normal for the year is five per cent below the trend-normal, and in Year II two per cent below.

The discrepancy is due entirely to the method of eliminating trend and seasonal, to the combination of seasonal percentages of average annual activity with ordinates of trend which are differing absolute values and not percentages of average annual activity. The difference in the percentages of discrepancy is due to the fact that the line of trend is not a logarithmic curve; on different annual segments of a given line which is rising or falling in value by other than a constant proportional amount from any one period to the next the several sub-annual points making up the annual segments will in every year, of course, bear different percentage relationships to the average values of the different segments. The size of the discrepancies is due to two factors: first, the steepness of the line of trend, for on a line of smaller slope the sub-annual points will have a smaller range of percentage variation from the averages of the annual segments than on a line of larger slope; and second, the intensity and order in time of the seasonal variation, for if the seasonal is of slight importance the trend-seasonal normal will necessarily be about the same as the trend-normal no matter how the measures of trend and seasonal are combined, or if the seasonal variation happens to be so distributed that its greatest fluctuations from the basic average are paired with trend values which are closest to the annual average trend activity the resulting trend-seasonal normal will be about the same under any method of combining trend and seasonal.

Logic requires a method which will permit the elimination of trend and seasonal without distortion of the normal. A method proposed to the Royal Statistical Society by Dr. E. C. Snow in his paper on "Trade Forecasting and Prices,"[1] in which he remarks incidentally that the logic of the usual method of allowing for seasonal is not clear, meets this requirement. The method consists of a simultaneous elimination of trend and seasonal through the process of grouping the items for each sub-annual period over the whole number of years involved and fitting lines of normal change to each such group, each item then being expressed as a percentage of the ordinate found for its year in its group. For an independent measure of seasonal a separate calculation is necessary. The same process has been suggested by Professor Crum as one method of allowing for progressive variation in seasonality.[2] These virtues of the method are offset by

[1] *Journal of the Royal Statistical Society*, May 1923, p. 332.
[2] This JOURNAL, March 1925, p. 48.

the disadvantage which Professor Crum mentions that, as in the case of the simple method of monthly means for measuring seasonal variation, it is liable to considerable inaccuracy if the data to which it is applied show much irregularity of variation.

A number of alternative methods suggest themselves. One general process is illustrated in Example B which may be applied regardless of how the seasonal index is computed and how the seasonal index may change from year to year. The same items, ordinates of trend, and seasonal index are assumed as in Example A. The essential point of the process lies in expressing each ordinate of trend as a percentage of the average trend activity for the year in which that ordinate is found, and combining these percentages with the corresponding figures of the seasonal index which are also, in the significance of their adjusted form, percentages of the average trend activity for the year, into a trend-seasonal index (Column f). The average of trend for a year is then distributed sub-annually by multiplying that average by each percentage figure of the trend-seasonal index, giving as the products the ordinates of normal (Column g) or the values which the variable would assume if trend and seasonal were the only variations involved. It will be observed that the average or total of the normal

EXAMPLE B

Time	Item	Ordinates of Trend	Trend Index	Seasonal Index	Trend-Seasonal Index	Ordinates of Normal	Per cent of Normal
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Year I							
1st quarter	12	10	40	120	60	15	80
2nd "	24	20	80	90	70	17.5	137
		— (25)					
3rd "	36	30	120	110	130	32.5	111
4th "	48	40	160	80	140	35	137
		100				100	
Year II							
1st quarter	60	50	77	120	97	63	95
2nd "	72	60	92	90	82	53.3	135
		— (65)					
3rd "	84	70	108	110	118	76.7	109
4th "	96	80	123	80	103	67	143
		260				260	

activity indicated is necessarily the same as the average or total which would have existed had there been no seasonal variation. The final step for the elimination of trend and seasonal, for securing percentages of normal (Column h), is simply the division of each original item by the corresponding ordinate of normal as calculated. This method requires only slightly more time for its application than does the usual

method. The only additional step which is at all time-consuming is the expression of each ordinate of trend as a percentage of the average trend activity in the respective years; as all the ordinates in any one year are divided by the same average figure, this computation is not at all laborious.

The method of combining the trend index with the seasonal index by adding together the two percentages and subtracting 100 from the sum is supported on the following grounds: the average of trend activity for the year may be considered the most probable normal value of the variable at any point in the year; the forces of trend and seasonal variation act against this most probable value to make it actually more or less at different points in the year; if the forces of trend act within the year to make the normal value at a particular point twenty per cent less than the average trend activity of that year, and if seasonal variation exerts its influence to make the normal value at the same point twenty per cent more than the average trend activity, the normal value at that point will be the same as the average trend activity, while if seasonal variation has tended to make the normal value only ten per cent more than the average trend activity, the net effect would have been to give a normal value at that point ten per cent less than the average trend activity.

The percentages of normal obtained here should be compared with the percentages of normal obtained by the customary method in Example A. The differences are a function of the factors cited in connection with Example A as causing the discrepancies there between the indicated normal and the original trend-normal, and also of the extent of variation away from normal of the original items of the data.

There is appended a chart to illustrate the extent of the distortion from the use of the customary method in one actual case. The data of the chart are Percentages of Normal of Refinery Stocks of Gasoline in the United States, quarterly (the average of three end-of-month actual items) from 1917 to 1925, the first quarter being in the months of December of the year preceding and January and February of the year stated, the second quarter March, April, and May, and so on. The original data reveal a large trend and pronounced seasonality, but the use of quarterly figures has reduced the latter somewhat. The line of trend which was fitted would not be satisfactory for a complete and continued analysis of this material, but its partial inappropriateness does not affect the point which is here of interest, namely the difference in results yielded by the two methods of securing a normal from which to measure deviations. The customary method gives results too high in the second and fourth quarters, too low in the first quarter. The

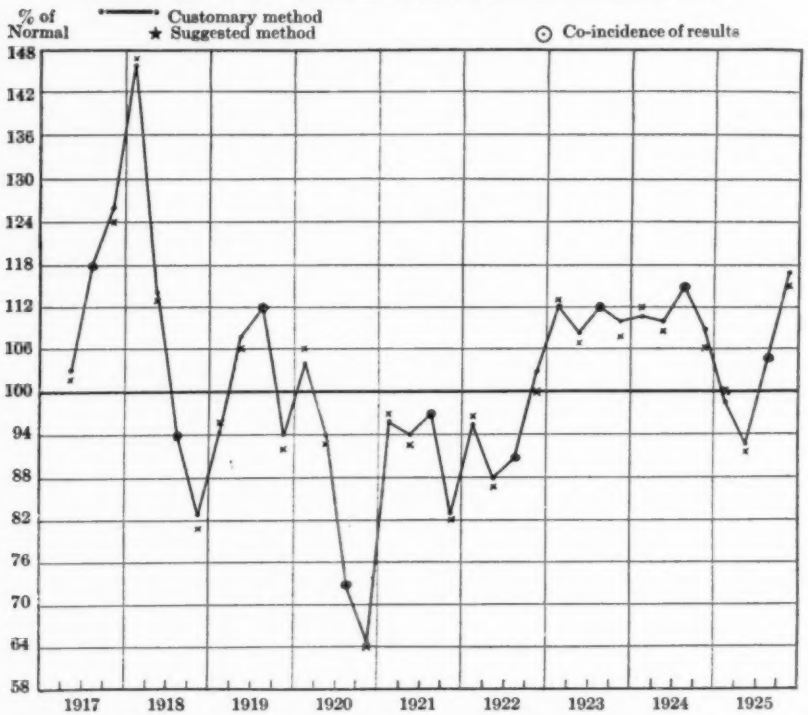
coincidence of results in the third quarter is due to the fact that for that quarter the seasonal percentage is 100. It is apparent that the differences in this case are not so great as to be a source of any important error in the general run of conclusions which might be drawn from the analysis of this material at the present time.

Except in those cases where the slope of the line of trend is very large, the seasonal variation very intense, and the data of a fairly wide range of fluctuation, the differences in the results from the two methods will not be great. In the majority of cases the differences will probably be of minor significance, for trends are rarely so steep that ordinates of trend depart very far from one hundred per cent of the average of trend for a year.

However good results the old method may yield in the great majority of cases, it yet remains true of that method that it is in theory

CHART I.—REFINERY STOCKS OF GASOLINE IN THE UNITED STATES, QUARTERLY,* 1917–1925

Source: Standard Daily Trade Service, Bankers Economic Service



* Average of 3 end-of-month items: First quarter—December, January, February; Second—March, April, May; Third—June, July, August; Fourth—September, October, November.

illogical and if the exceptional cases appear and call for analysis it is then capable of giving only a misleading representation of those fluctuations which we ascribe to a business cycle. With the possibility of more homogeneous data and the certainty of greater refinement in related statistical methods in the future, it would seem wise to secure as precise and rigorous a method as possible for the determination of the trend-seasonal normal in time series. This consideration has been the only excuse for the criticism and suggestions here presented.

NOTES

A PARTICULAR CASE OF NON-REPRESENTATIVE SAMPLING

By G. Pietra, *University of Ferrara, Italy*

A recent paper by Warren S. Thompson, "Size of Families from which College Students Come," [1] as a contribution to those researches which have attempted to ascertain the number of children born to parents in a particular group, offers me the opportunity to reconsider an aspect of the method of securing sample data in order that sampling may correctly represent the entire field.

To establish the average size of students' families Mr. Thompson requested a certain number of students to state the number of their brothers and sisters.

Now, it is to be remarked that if f_i shows the number of children which belong to the i^{th} of n families the average number of children will be:

$$N = \frac{\sum_1^n f_i}{n}$$

but if we choose at random a child of the above-said n families, since the probability of choosing this child among those which belong to the i^{th} family is proportional to f_i, the probable number of children born to his parents will be instead:

$$N_1 = \frac{\sum_1^n f_i^2}{\sum_1^n f_i}$$

and it is easy to prove that

$$N_1 > N.$$

Therefore, in Mr. Thompson's collected data there is a selection, which makes the average size of students' families larger than that which would result in the totality of the field.

Let all families present the same probability of having in each child

[1] This Journal, December, 1925, p. 481.

a student, then, in this case, the difficulty can be easily removed by dividing the frequency, that is, the number of families, by the intensity of the character, that is, the number of children.

The hypothesis, however, that the number of students is proportional to the total children of the same family may hardly be accepted, since it is well known that the probability of having a student is greater in the case of families of higher social degree. Families of higher social degree, however, offer a low marriage fertility. In this case to adjust the sample data it will be necessary to ascertain the relation of the number of students to that of all the children of each family; but it is not my intention to go further in this argument.

I wish, instead, only to note that methodological objections, analogous to those I have here briefly stated with reference to Mr. Thompson's paper, were directed sometime ago against other statisticians by Professor Gini and Professor Weinberg.[1] Thus, Professor Gini demonstrates that no conclusions may be drawn from the researches of Steinmetz [2] for Holland as to lower prolificness in prominent people and points out that we may question the results of Pearson, which would show that prolificness of degenerate families is higher than that of normal families; [3] while a correct use of Gini's method could prove, for many cases, the contrary, in accordance with Weinberg's results.

These lines were already written when my attention was called to an article by Professor Lenz in the *Eugenical News* of last January. Professor Lenz makes about the same objections that I have raised here, in regard to a work of Professor L. M. Terman, *Genetic Studies of Genius*, in order to demonstrate that those families from which the children studied came represent a selection of those with more than the average number of children. I think, however, that my remarks may not be entirely useless.

In fact, it is to be noticed that Professor Lenz correctly cleared up by a simple example the difficulty presented by Terman's sample data and analogous to that of Mr. Thompson. He properly rectified it in the hypothesis (which, however, as we have already noticed, can be only admitted approximately) that the probability for each family to be included in the random sample is proportional to the number of its

[1] C. Gini: "I fattori demografici dell'evoluzione delle nazioni," Torino, Bocca, 1912; and "Contributi Statistici ai problemi dell'Eugenics," *Rivista Italiana di Sociologia*, Roma, maggio-agosto 1912.

This second paper reproduces, with some additions and illustrations, a lecture made before the First International Congress of Eugenics in London, 1912. See "Contributions of Demography to Eugenics," "Problems in Eugenics," Vol. II.

Also Weinberg, "Die rassenhygienisch Bedeutung der Fruchtbarkeit," *Archiv für Rassen-u. Gesellschafts-Biologie*, 1910, Heft 6; 1911, Heft 1.

[2] "Der Nachwuchs der Begabten," *Zeitschrift für Socialwissenschaft*, 1904, Heft 1.

[3] *The Scope and Importance to the State of the Science of National Eugenics*, Eugenics Laboratory Lecture, Series I. Dulan and Co., London, 1911.

children and then suggests dividing the number of families by the number of children, in order to get the same composition of the material as in the totality of all fertile marriages. Examining a little further the size of families of parents from which gifted children come, Professor Lenz considers the decline in fertility of gifted families. He shows thus an evident contradiction of the hypothesis which he adopted in making his calculations.

It is to be noticed that with the hypothesis that all families have the same probability of having in each child a student, or any kind of gifted or degenerate child, the composition of the parents' families is larger than that of the children although no decline in fertility exists.

This is a natural consequence of the fact that if the occurrence of a character is only probable and not certain, it will be more likely to occur among those families which are more numerous although the probability of the occurrence of the character in each child is independent of the number.

THE SIZE OF FAMILIES

A DISCUSSION OF SAMPLING

By Elbridge Sibley

Professor Pietra sets forth a point of methodology which the writer had raised less nicely at about the same time. It seems to be possible to carry the same line of reasoning further, to apply to the problem of computing not only the average size of parents' families from a sample of the filial generation, but also the average size of grandparents' families. The latter problem is involved in such studies as the recent one of Professor Thompson[1] cited by Professor Pietra, and that of Baber and Ross,[2] both of which show astonishing differences in the average sizes of families in the two generations.

It may be stated in general that the probability of a given person's being represented by one descendant in a random sample of a later generation is in exact proportion to the number of descendants that that person has in the generation from which the sample is drawn. It is further apparent that the number of descendants that a person has will equal the number of his children, multiplied by the average number of (grand-) children born to each of them, and so on.

We shall consider in detail only the case of grandparents and grandchildren, although the method would be applicable to any number of generations.

If in a given population there are f_n families having n children each, and the nf_n persons in the second generation have an average of p_n children apiece (it being assumed that there is no intermarriage), the third generation will number nf_np_n. Now if we denote the number of individuals in the third generation by q_n,

$$q_n = nf_np_n \tag{1}$$

whence

$$f_n = q_n/np_n. \tag{2}$$

The average size of the first generation families,

$$N = \frac{\Sigma nf_n}{\Sigma f_n}. \tag{3}$$

[1] This Journal, December, 1925.

[2] University of Wisconsin Studies in Social Science, No. 10, 1924.

Substituting the value of f_n from equation (2) gives

$$\overline{N} = \frac{\Sigma n(q_n/np_n)}{\Sigma(q_n/np_n)} = \frac{\Sigma(q_n/p_n)}{\Sigma(q_n/np_n)}. \tag{4}$$

The mean sizes of grandparental families given in the studies cited have evidently been computed without regard to the weighting of the sample, from the equation

$$\overline{N}_1 = \frac{\Sigma nq_n}{\Sigma q_n}. \tag{5}$$

From equations (4) and (5) it is seen that the relationship between the true mean size of the grandparents' families ($\overline{N}$) and the computed value ($\overline{N}_1$) depends upon (a) the relation between n and f_n, and (b) the relation between n and p_n. Professor Pietra takes account of the first factor by using the formula

$$\overline{N}_2 = \frac{\Sigma q_n}{\Sigma(q_n/n)} \tag{6}$$

and he points out that (unless all the families in the population be of the same size)

$$\overline{N}_1 > \overline{N}_2. \tag{7}$$

We may now consider the relationship of $\overline{N}$ and $\overline{N}_2$: There are three possibilities:

(1) If p_n is the same for all values of n, equations (4) and (6) are identical. This would be the case if there were no biological or social inheritance of fertility.

(2) If p_n is negatively correlated with n—a most unlikely assumption —$\overline{N}$ will be greater than $\overline{N}_2$.

(3) If p_n is positively correlated with n—as is most probably the case—$\overline{N}$ will be less than $\overline{N}_2$.

Our formula (4) could be applied as it stands only if data were available not only on the fertile but also on the infertile families in the second generation. But since it is not the absolute but the relative magnitudes of the several values of p_n that affect the result, and since in populations which have been studied there seems to be a fairly high negative correlation between average size of fertile families and proportion of infertile families, it would seem legitimate to take as a first approximation to the relative values of p_n the average sizes of fertile families only, i.e., the families of the parents of the q_n individuals in the sample. It seems likely that such a procedure would indicate a reduction in the apparently rapid rate of decline in family size shown by the studies which we have mentioned.

OCCUPATIONAL GROUPS IN THE UNITED STATES, 1820–1920

By P. K. Whelpton, *Scripps Foundation*

A note on statistics of persons employed at census dates since 1820 at first sight seems superfluous, in view of the vast amount of space given to data on occupations in various census reports. But those who have desired comparable series covering the past century on such matters as the number of people engaged in agriculture, and who have tried to find them in census publications are almost sure to have run into several difficulties. For example, the classification of occupations into general divisions has been changed several times in this period, fishermen being listed under manufacturing at one time and under agriculture at another. More serious than this, certain census reports have included under domestic and personal service several hundred thousand laborers most of whom were working on farms, in factories, or for railroads, but whose occupations were not described in sufficient detail by the enumerators or marshalls to enable proper classification at the central office. Again, some groups have been omitted entirely as far as the occupation statistics of certain censuses are concerned, no data being gathered for slaves in 1860, nor for slaves or free females in 1850. Considering the conditions under which many of the early censuses were taken, the wonder probably is that no greater disparities occurred. But to use the occupation figures as they stand may cause rather serious errors in some instances.

To attempt to straighten out some of these difficulties and secure comparable series of the number of workers in agriculture, manufacturing, and other important groups during the past century is the purpose in mind here. The additions and changes that seem necessary require many estimates, of course. While these are manifestly subject to error themselves, it is believed that they have some reasonable basis, and that they lead to results which more nearly represent actual conditions than do the data as they now appear in census publications.

A summary of the number of persons engaged in each general division of occupations is presented in Table 1, Section A. The proportion they constitute of the persons 10 years of age and over, and also of the employed persons is brought out in Sections B and C of the same table. The striking changes during the century are the ten-fold increases in the relative importance of workers in trade and transportation, and in

mining. The proportion in manufacturing was about two and a half times as high in 1920 as in 1820, while professional service had almost doubled. Agriculture, the first numerically up to 1920 when it dropped to second place, fell considerably in relative importance, standing only about forty per cent as high in 1920 as in 1820. Domestic and personal service, and fishing showed little change during the period. Lumbering, though unimportant numerically, increased about threefold in relative importance.

Differences between division totals as they appear here and in the census are rather marked in several instances. For example, domestic and personal service in 1900 amounts to 3,076,000 [1] compared to 5,580,657 in the census,[2] a reduction of 2,504,657 or 44.9 per cent due to reclassifying laborers not specified. This process adds 1,411,000 to the number engaged in manufacturing that year,[3] an increase of 21.9 per cent. Similarly, 662,000 laborers are added to agriculture in 1870,[4] an increase of 10.6 per cent. More obvious is the effect of allowing for slaves and free females 10 years of age and over and free males 10 to 15 years of age occupied in 1850. The new total here is 7,697,000 compared with 5,329,506,[5] an increase of 44.4 per cent.

Such revisions naturally carry over into studies based on census material. For example, changes in the efficiency of agriculture, i. e., the output per worker, have sometimes been measured by comparing the ratio of the total population to the agricultural workers. Using the number of agricultural workers as given in the census,[6] this ratio would be 6.51 in 1870 and 7.41 in 1910, a relative gain in efficiency of 13.8 per cent. Using the revised data,[7] the ratio would be 5.77 in 1870 and 7.86 in 1900, indicating an increase in efficiency of 36.2 per cent, more than two and a half times as great as the uncorrected result, and manifestly representing much more accurately what has actually happened.

The general divisions of occupations are dealt with separately in Tables 2 to 5, and the method of compilation shown in detail. Two divisions introduced in the 1910 Census for the first time, clerical and public service, were not used, as there seemed to be inadequate data for carrying them back through earlier census reports. Instead, they were

[1] Table 1, Section A.
[2] 13th Census, Vol. IV, p. 54.
[3] Table 5, Column B.
[4] Table 4, Note m.
[5] 12th Census, Special Report Occupations, p. 411.
[6] Population, 14th Census, Vol. I, p. 14. Agricultural workers, 1870, 12th Census, Special Report Occupations, p. L, division total less lumbermen and woodchoppers; 1910, 13th Census, Vol. IV, p. 54, same items.
[7] Population, 14th Census, Vol. I, p. 14, revised data for 1870. Agricultural workers, Table I, Section A.

distributed for 1910 and 1920 among the remaining divisions according to the old classification. Similarly, trade and transportation, treated as one group before 1910 but two groups in 1910 and 1920, were combined in these latter years as it seemed impractical to carry the separation back.

Brief mention of some of the other changes and additions made may be desirable. It was estimated that an over-enumeration of 713,000 agricultural workers occurred in 1910 and an under-enumeration of 240,000 in 1920. The former number appears to have a fairly adequate base,[1] but the margin of error in the latter may be wide.[2] However, a change of 25 per cent in this estimate would only change the division total one-half of one per cent.

A more serious matter was the allocation to the proper divisions of the laborers not specified included in domestic and personal service under the old classification who amounted to as many as 2,629,000 in 1900. In 1910, the new classification did away with this group,[3] but for the sake of comparison with earlier censuses, a second tabulation was made based on the old classification.[4] In reconciling these two classifications it became apparent that the bulk of these laborers in 1910 belonged in manufacturing and in trade and transportation, with a few in domestic and personal. In 1890, 1880, and 1870, however, a part of these laborers not specified were actually engaged in agriculture.[5]

In 1900, it is not so clear whether agricultural laborers were included or not. An attempt was made to exclude them when editing the returns,[6] but in spite of this, the proportion of occupied persons reported as laborers not specified is found to average 7.5 per cent in 5 industrial states,[7] 7.0 per cent in 5 southern agricultural states,[8] and 7.3 per cent in 5 northern agricultural states.[9] The uniformity of these per cents seems to indicate that a considerable part of the laborers not specified in 1900 belonged in agriculture.

Attempts were made to distribute laborers not specified among the divisions concerned in proportion to the size of the divisions, but the results did not seem satisfactory. It was finally decided that a more accurate series for agriculture could be secured by multiplying the total number of employed persons by the per cent engaged in agriculture, this latter being known for 1820, 1840, 1910, and 1920, and

[1] Table 4, Note i.
[2] Table 4, Note d.
[3] 13th Census, Vol. IV, p. 53.
[4] *Op. cit.*, p. 54.
[5] 11th Census, Population, part II, p. LXXXVIII, and 10th Census, Population, p. 744.
[6] 12th Census, Special Report Occupations, pp. XXV and XXVI.
[7] Massachusetts, Rhode Island, Connecticut, New York, and New Jersey.
[8] North Carolina, South Carolina, Georgia, Alabama, and Mississippi.
[9] Minnesota, Iowa, Missouri, Nebraska, and Kansas.

estimated for the remaining years by means of the per cent of the total population living in rural territory. The details of this method and the number of laborers not specified allocated to agriculture are shown in Table 4. The domestic and personal service division was then increased by a small percentage to allow for laborers belonging to it,[1] and the balance was divided between manufacturing, and trade and transportation, based on the 1910 data previously mentioned.[2]

In certain earlier censuses no occupation data were given for 5 divisions, domestic and personal service, professional service, mining, lumbering, and fishing. In these cases the per cent of the population 10 years of age and over engaged in each division was estimated by continuing back the trend for the known years. Applying these per cents to the number of persons 10 years of age and over gave the estimated number of persons engaged in each occupational division.[3]

For 1870, it was considered necessary to increase the number of occupied persons in every division except lumbering because of the estimated number of persons omitted from the Ninth Census in the southern states.[4]

No occupation statistics for slaves, or for free persons 10 to 15 years of age were secured in the 1860 Census, nor for these groups or other free females in the 1850 Census. After studying the material it appeared that these omissions were comparatively small in the mining, fishing, and lumbering divisions, and estimates of the numbers concerned were made, based on 1870 age and sex ratios, and 1890 color ratios with allowance for greater concentration of Negroes in agriculture under conditions of slavery.[5] With the professional division it appeared that practically no slaves or free persons under 16 years of age were concerned and hence no adjustment necessary for 1860. For 1850, the number of free females over 15 omitted was estimated, based on the 1860 data.[6] A larger adjustment seemed necessary in the domestic and personal service division and was finally made by assuming the proportion of the population 10 years of age and over engaged in this division to be the same as the average of the 1870 and 1880 proportions. The method used for agriculture has previously been described.

With manufacturing, and trade and transportation, the two remaining divisions, a plan similar to that just mentioned for domestic and personal service could have been followed, namely estimating the per cent of the population 10 years of age and over in these divisions in

[1] Table 3, Column G.
[2] Table 5, Columns B and F.
[3] Table 2, Notes q, r, and s. Table 3, Notes u and w.
[4] Table 4, Notes n and o.
[5] Table 2, Notes j and o.
[6] Table 3, Note v.

1850 and 1860 based on the known per cents in other years. An alternative method consisted in estimating the per cent of the population 10 years of age and over employed in these years based on known per cents in other years, calculating the total number employed, and from this the number in manufacturing and trade and transportation by subtracting the number employed in the other six divisions. The result was divided between manufacturing, and trade and transportation according to known proportions in 1820, 1840, and 1870 and later years.[1] These two methods gave very similar results: the latter was finally used.

Although these revisions are believed to be more nearly comparable throughout the period, and to represent conditions more accurately than the published census figures on which they are based, there is little doubt but that still better results could be secured by referring to some of the original schedules, and to certain unpublished tabulations which probably exist. The Census Bureau itself is the only agency in position to follow this plan of work. It is to be hoped that enough interest in such revisions may be aroused that the Bureau will feel justified in making the attempt and pushing it to completion.

TABLE 1

NUMBER OF PERSONS ENGAGED IN GAINFUL OCCUPATIONS, CLASSIFIED BY GENERAL DIVISIONS OF OCCUPATIONS (a)

A. NUMBER OF PERSONS (THOUSANDS)

Year	All Occupations (b)	Agriculture (c)	Manufacturing and Mechanical Pursuits (d)	Trade and Transportation (e)	Domestic and Personal Service (f)	Professional Service (g)	Mining (h)	Lumbering (i)	Fishing (j)
1920.....	41,854	10,923	12,819	10,433	4,207	2,112	1,090	217	53
1910.....	37,454	11,704	10,629	7,989	4,219	1,706	965	174	68
1900.....	29,073	10,699	7,854	5,444	3,076	1,213	597	121	69
1890.....	23,319	9,770	5,969	3,650	2,403	947	408	112	60
1880.....	17,390	8,505	4,185	2,122	1,622	602	264	49	41
1870.....	12,927	6,904	2,746	1,347	1,318	379	176	29	28
1860.....	10,531	6,287	1,932	783	1,003	305	168	23	30
1850.....	7,697	4,965	1,261	415	735	207	89	13	12
1840.....	5,420	3,718	792	207	520	147	15	9	12
1830.....	3,940	2,773	524	122	386	109	11	6	9
1820.....	2,881	2,071	350	72	288	81	8	5	6

B. PER CENT OF POPULATION 10 YEARS OF AGE AND OVER (k)

1920.....	50.59	13.20	15.49	12.62	5.08	2.55	1.32	.26	.06
1910.....	52.32	16.35	14.85	11.16	5.89	2.38	1.35	.24	.10
1900.....	50.17	18.46	13.55	9.39	5.31	2.09	1.03	.21	.12
1890.....	49.18	20.61	12.59	7.70	5.07	2.00	.86	.24	.13
1880.....	47.31	23.14	11.38	5.78	4.41	1.64	.72	.13	.11
1870.....	44.39	23.71	9.43	4.63	4.53	1.30	.60	.10	.10
1860.....	46.95	28.03	8.61	3.49	4.47	1.36	.75	.10	.13
1850.....	46.78	30.18	7.67	2.52	4.47	1.26	.54	.08	.07
1840.....	46.61	31.97	6.81	1.78	4.47	1.26	.13	.08	.10
1830.....	45.66	32.13	6.07	1.41	4.47	1.26	.13	.07	.10
1820.....	44.71	32.14	5.43	1.12	4.47	1.26	.13	.07	.10

[1] Table 5, Note q.

C. PER CENT OF OCCUPIED PERSONS (l)

1920.....	100.0	26.10	30.63	24.93	10.05	5.05	2.60	.52	.13
1910.....	100.0	31.25	28.38	21.33	11.26	4.55	2.58	.46	.18
1900.....	100.0	36.80	27.01	18.73	10.58	4.17	2.05	.42	.24
1890.....	100.0	41.89	25.60	15.65	10.31	4.06	1.75	.48	.26
1880.....	100.0	48.90	24.07	12.20	9.33	3.46	1.52	.28	.24
1870.....	100.0	53.41	21.24	10.42	10.20	2.93	1.36	.22	.22
1860.....	100.0	59.70	18.35	7.44	9.52	2.90	1.60	.22	.28
1850.....	100.0	64.51	16.38	5.39	9.55	2.69	1.16	.17	.16
1840.....	100.0	68.60	14.61	3.82	9.59	2.71	.28	.17	.22
1830.....	100.0	70.38	13.30	3.10	9.80	2.77	.28	.15	.23
1820.....	100.0	71.88	12.15	2.50	10.00	2.81	.28	.17	.21

(a) Includes persons 10 years of age and over from 1850 to 1920; and persons "engaged in actual occupations" (12th Census, Sp. Rpt. Occupations, p. XXIX) from 1820 to 1840.
(b) Table 4, Column C.
(c) Table 4, Column G.
(d) Table 5, Column C.
(e) Table 5, Column G.
(f) Table 3, Column H.
(g) Table 3, Column D.
(h) Table 2, Column A.
(i) Table 2; Column C.
(j) Table 2, Column E.
(k) Each column of Section A divided by Table 4, Column A.
(l) Each column of Section A divided by Table 4, Column C.

TABLE 2

PERSONS ENGAGED IN MINING, LUMBERING, AND FISHING AND OYSTERING; NUMBER (IN THOUSANDS), AND PER CENT OF POPULATION 10 YEARS OF AGE AND OVER

Year	Mining		Lumbering		Fishing and Oystering	
	A Number	B Per Cent (a)	C Number	D Per Cent (a)	E Number	F Per Cent (a)
1920.....	1,090 (b)	1.32	217 (e)	.26	53 (d)	.06
1910.....	965 (e)	1.35	174 (e)	.24	68 (d)	.10
1900.....	597 (f)	1.03	121 (g)	.21	69 (h)	.12
1890.....	408 (f)	.86	112 (g)	.24	60 (h)	.13
1880.....	264 (f)	.72	49 (g)	.13	41 (h)	.11
1870 (i).....	176 (f)	.60	29 (g)	.10	28 (h)	.10
1860 (j).....	168 (k)	.75	23 (m)	.10	30 (n)	.13
1850 (o).....	89 (k)	.54	13 (m)	.08	12 (n)	.07
1840.....	15 (p)	.13	9	.08 (q)	12	.10 (r)
1830.....	11	.13 (s)	6	.07 (q)	9	.10 (r)
1820.....	8	.13 (s)	5	.07 (q)	6	.10 (r)

(a) Preceding column divided by Table 4, Column A.
(b) 14th Census, Vol. IV, p. 35, line 51.
(c) *Ibid.*, line 23, 35, and 40.
(d) *Ibid.*, line 22.
(e) Same as Note b, based on new classification. Based on old classification, 13th Census, Vol. IV, p. 55, oil well employees, salt well and works employees, and miners and quarrymen; plus 2.38%, the per cent by which the division total for 1910 according to the old classification is lower than the total according to the new classification. It is assumed that this represents the operators, officials, and managers, etc., listed separately in 1910 and 1920 for the first time, and applies to each census previous to 1910.
(f) 12th Census, Sp. Rpt. Occupations, p. LI, oil well employees, salt works employees, and miners and quarrymen; plus 2.38%.
(g) *Op. cit.*, p. L, lumbermen, raftsmen and woodchoppers; plus 12%, the average of the per cents which certain occupations listed under this division in 1910 and 1920, but not in earlier censuses, were of the balance of the division in those years.
(h) *Op. cit.*, p. LI, fishermen and oystermen.
(i) Column A and Column E each increased by 1,000 to allow for faulty enumeration in the Southern States. (See Table 4, Note p.)
(j) Column A increased by 5,000, Column C by 1,000, and Column E by 3,000 to allow for slaves, and free persons 10 to 15 years of age, no data on occupations being obtained for these groups by the census.
(k) 12th Sp. Census, Rpt. Occupations, p. LVII and LVIII, oil well and works employees, saltmakers, and miners and quarrymen; plus 2.38%.
(m) *Op. cit.*, lumbermen, raftsmen, and woodchoppers; plus 12%.
(n) *Op. cit.*, fishermen and oystermen.
(o) Column A increased by 3,000, Column C by 600, and Column E by 1,000 to allow for slaves and free females, and free males 10 to 15 years of age. (See Note j.)
(p) 6th Census, p. 475.
(q) Estimated along a smooth curve.
(r) Estimated to be the same as the average of 1850 and 1860.
(s) Estimated to be the same as 1840.

TABLE 3

PERSONS ENGAGED IN PROFESSIONAL SERVICE AND IN DOMESTIC AND PERSONAL SERVICE; NUMBER (IN THOUSANDS), AND PER CENT OF POPULATION 10 YEARS OF AGE AND OVER

Year	Professional Service					Domestic and Personal Service			
	A Total Number Reported	B No. of Photographers	C Trans. to Domestic and Personal Service (a)	D Adjusted Total	E Per Cent (b)	F Total Number Reported	G No. of Domestic and Personal Laborers (c)	H Adjusted Total	I Per Cent (b)
1920.....	2,112 (d)	..	..	2,112 (d)	2.55	4,207 (e)	..	4,207 (e)	5.08
1910 (f)...	1,706 (d)	..	..	1,706 (d)	2.38	4,219 (e)	..	4,219 (e)	5.89
1910 (g)...	1,705 (h)	32 (i)	31	1,706 (j)	2.38	4,044 (k)	144	4,219 (m)	5.89
1900.....	1,208 (n)	27 (o)	22	1,213 (j)	2.09	2,951 (p)	103	3,076 (m)	5.31
1890.....	944 (q)	20 (o)	17	947 (j)	2.00	2,307 (p)	78	2,403 (m)	5.07
1880.....	603 (q)	10 (o)	11	602 (j)	1.64	1,560 (p)	52	1,622 (m)	4.41
1870 (r)...	378 (q)	8 (o)	7	379 (j)	1.30	1,272 (p)	39	1,318 (m)	4.53
1860.....	308 (s)	3 (t)	6	305 (j)	1.36	.....	..	1,003	4.47 (u)
1850.....	210 (v)	1 (t)	4	207 (j)	1.26	.....	..	735	4.47 (u)
1840.....	.....	..	..	147	1.26 (w)	.....	..	520	4.47 (u)
1830.....	.....	..	..	109	1.26 (w)	.....	..	386	4.47 (u)
1820.....	.....	..	..	81	1.26 (w)	.....	..	288	4.47 (u)

(a) After transferring all possible items in an endeavor to make the old classification for 1910 equal the new, an excess of 31,000 or 1.78% remains, which apparently should be transferred to the domestic and personal service division, Column H. Similar decreases of 1.78% are assumed to apply to earlier censuses.

(b) Preceding column divided by Table 4, Column A.

(c) 14th Census, Vol. IV, p. 41, line 72; plus p. 43, line 13; plus 24,000 (probably laborers not specified), an amount necessary to reconcile the two classifications after making all possible transfers. Similar proportional increases are assumed to apply to earlier censuses.

(d) *Op. cit.*, p. 42, lines 11 and 1; less lines 41, and 59 to 64; plus p. 41, lines 78, 79, 80.

(e) *Op. cit.*, p. 43, line 1; plus p. 41, lines 70, 71, 72, 76, 77; plus p. 42, lines 5, 6, 7, 41, and 59 to 64.

(f) Based on new classification.

(g) Based on old classification.

(h) 13th Census, Vol. IV, p. 54, division total; less electricians (14th Census, Vol. IV, p. 36, line 30).

(i) *Op. cit.*, p. 56.

(j) Column A, plus Column B, less Column C.

(k) 13th Census, Vol. IV, p. 54, division total, less laborers not specified.

(m) Total of Columns C, F, and G.

(n) 12th Census, Sp. Rpt. Occupations, p. L, division total; less electricians, p. XXXIV.

(o) *Op. cit.*, p. LII.

(p) *Op. cit.*, p. L, division total, less laborers not specified.

(q) *Op. cit.*, p. L.

(r) Column A has been increased by 7,000 and Column F by 41,000 to allow for faulty enumeration in the Southern States. (See Table 4, Note o.)

(s) 12th Census, Sp. Rpt. Occupations, p. LIV, division total; less stationary engineers (estimated to be 22,000).

(t) *Op. cit.*, p. LXII.

(u) Estimated to be the same as the average of 1870 and 1880.

(v) *Op. cit.*, p. LIV, division total; plus 51,000 women, estimated on the basis that the occupations in this division probably filled chiefly by men have the same relation to the total in 1850 as in 1860.

(w) Estimated to be the same as in 1850.

TABLE 4

PERSONS ENGAGED IN AGRICULTURE; NUMBER (IN THOUSANDS), AND PER CENT OF POPULATION 10 YEARS OF AGE AND OVER

Year	A Number of Persons 10 Years of Age and Over	B Per Cent Engaged in Gainful Occupations	C Number of Persons Engaged in Gainful Occupations	D Per Cent of Occupied Persons Engaged in Agriculture	E Per Cent of Total Population in Places Under 2500	F Ratio of D to E	G Number of Persons Engaged in Agriculture	H Per Cent of Persons 10 Years of Age and Over in Agriculture (a)
1920.....	82,739 (b)	50.59 (c)	41,854 (d)	26.1 (e)	48.6 (f)	.537 (g)	10,923 (h)	13.20
1910.....	71,580 (b)	52.32 (c)	37,454 (i)	31.2 (e)	54.2 (f)	.576 (g)	11,704 (h)	16.35
1900.....	57,950 (b)	50.17 (c)	29,073 (b)	36.8 (j)	60.0 (f)	.613 (k)	10,699 (m)	18.46
1890.....	47,414 (b)	49.18 (c)	23,319 (b)	41.9 (j)	64.6 (f)	.649 (k)	9,770 (m)	20.61
1880.....	36,762 (b)	47.31 (c)	17,390 (b)	48.9 (j)	71.4 (f)	.685 (k)	8,505 (m)	23.14
1870.....	29,124 (n)	44.39 (c)	12,927 (o)	53.4 (j)	74.1 (p)	.720 (k)	6,904 (m)	23.71
1860.....	22,430 (q)	46.95 (r)	10,531 (s)	59.7 (j)	79.6 (p)	.750 (k)	6,287 (m)	28.03
1850.....	16,453 (t)	46.78 (r)	7,697 (s)	64.5 (j)	84.0 (p)	.768 (k)	4,965 (m)	30.18
1840.....	11,628 (u)	46.61 (c)	5,420 (v)	68.6 (e)	88.8 (w)	.773 (g)	3,718 (u)	31.97
1830.....	8,630 (x)	45.66 (y)	3,940 (s)	70.4 (j)	91.2 (p)	.772 (k)	2,773 (m)	32.13
1820.....	6,444 (z)	44.71 (c)	2,881 (aa)	71.9 (e)	93.4 (p)	.770 (g)	2,071 (bb)	32.14

(a) Column G divided by Column A.
(b) 14th Census, Vol. IV, p. 33.
(c) Column C divided by Column A.
(d) 14th Census, Vol. IV, p. 35; plus 240,000, to allow for the estimated reduction in the number of farm laborers due to changing the census date from April 15 to January 1. Holmes estimates a reduction of 1,200,000 laborers actually at work (The Agricultural Situation, Vol. IV, No. 6), but it is assumed that only 20% of them would fail to return agriculture as their occupation, even though not employed in it at the time.
(e) Column G divided by Column C.
(f) 14th Census, Vol. I, p. 43.
(g) Column D divided by Column E.
(h) 14th Census, Vol. IV, p. 35, lines 4, 9, 15, 24, 30, and 43; plus 240,000 in 1920 (see Note d); less 713,000 in 1910 (see Note i).
(i) *Op. cit.*, p. 33; less 713,000 to allow for an over-enumeration in agriculture of 468,000 females (13th Census, Vol. IV, p. 28) and 245,000 males 10 to 15 years of age. This last is obtained by reducing the per cent of the age group employed in agriculture to 14.2%, which is an average of the 1900 per cent (17.7% weighted by 2) and the 1920 per cent (7.3%). The 1920 per cent is considered less accurate than that of 1900 (Note d), hence is given less weight.
(j) Column E multiplied by Column F.
(k) Interpolated along a smooth curve.
(m) Column C multiplied by Column D. It is estimated that Column G includes 438,000 laborers not specified in 1900, 733,000 in 1890, 840,000 in 1880, and 662,000 in 1870. These amounts are calculated as the excess of Column G over the following:—for 1900, 1890, and 1880, total agricultural pursuits (13th Census, Vol. IV, p. 54) less Table 2, Column C; for 1870, total agricultural pursuits (12th Census, Sp. Rpt. Occupations, p. L) plus 323,000 (estimated number of the 423,000 in Note o engaged in agriculture), less Table 2, Column C.
(n) 9th Census, Vol. I, p. 674, plus 895,000, calculated by applying the per cent of the population 10 years of age and over in 12 Southern States (71.05%) to 1,260,000, the estimated number of persons omitted from the Census, in the Southern States (14th Census, Vol. I, p. 14).
(o) 12th Census, Sp. Rpt. Occupations, p. L, plus 423,000, calculated by applying the per cent of the population 10 years of age and over engaged in gainful occupations in 12 Southern States (47.26%) to 895,000 (see Note n). The 423,000 are distributed to various divisions based on the distribution of those enumerated for these States. (See Table 2, Note i, Table 3, Note r, and Table 5, Note p.)
(p) Estimated, based on a smooth curve of the ratio of Column E to the per cent of the population in places under 8,000 for 1840, and 1880 to 1920.
(q) 8th Census, Vol. I, p. 596.
(r) Interpolated on a straight line between 1840 and 1880, the 1870 figure not being used as it very probably is affected by abnormal conditions which followed the Civil War.
(s) Column A multiplied by Column B.
(t) 7th Census, p. XLII to XLIV.
(u) 6th Census, p. 474 to 476.
(v) *Op. cit.*, p. 475; plus Table 2, Columns C and E; plus Table 3, Columns D and H.
(w) Calculated from 1840 Census, excludes 173 places of 2,500 or more with aggregate population of 1,911,000.
(x) 5th Census, p. 162 and 163.
(y) Interpolated on a straight line between 1820 and 1840.
(z) 4th Census, summary table, white persons, 5,236,000; slaves and free colored estimated to be 1,208,000 based on age ratios in the 10th Census, Vol. Population, p. 548 to 551, and 7th Census, p. XLIII and XLIV.
(aa) 4th Census, summary tables; plus Table 2, Columns A, C, and E; plus Table 3, Columns D and H.
(bb) 4th Census, summary tables.

TABLE 5

PERSONS ENGAGED IN MANUFACTURING AND MECHANICAL PURSUITS, AND IN TRADE AND TRANSPORTATION; NUMBER (IN THOUSANDS), AND PER CENT OF POPULATION 10 YEARS OF AGE AND OVER

Year	Manufacturing and Mechanical Pursuits				Trade and Transportation			
	A Number Reported	B Laborers Added	C Total	D Per Cent (a)	E Number Reported	F Laborers Added	G Total	H Per Cent (a)
1920.....	12,819 (b)	....	12,819 (b)	15.49	10,433 (c)	...	10,433 (c)	12.62
1910 (d)..	10,629 (b)	....	10,629 (b)	14.85	7,979 (c)	...	7,989 (c)	11.16
1910 (e)...	9,863 (f)	766 (g)	10,629 (h)	14.85	7,606 (i)	383 (j)	7,989 (k)	11.16
1900.....	6,443 (m)	1,411 (n)	7,854 (h)	13.55	4,767 (o)	677 (n)	5,444 (k)	9.39
1890.....	5,191 (m)	778 (n)	5,969 (h)	12.59	3,326 (o)	324 (n)	3,650 (k)	7.70
1880.....	3,469 (m)	716 (n)	4,185 (h)	11.38	1,872 (o)	251 (n)	2,122 (k)	5.78
1870 (p)..	2,496 (m)	250 (n)	2,746 (h)	9.43	1,264 (o)	82 (n)	1,347 (k)	4.63
1860.....	.....	....	1,932 (q)	8.61	.....	...	783 (q)	3.49
1850.....	.....	....	1,261 (q)	7.67	.....	...	415 (q)	2.52
1840.....	.....	....	792 (r)	6.81	.....	...	207 (r)	1.78
1830.....	.....	....	524 (q)	6.07	.....	...	122 (q)	1.41
1820.....	.....	....	350 (t)	5.43	.....	...	72 (t)	1.12

(a) Preceding column divided by Table 4, Column A.
(b) 14th Census, Vol. IV, p. 35, line 71.
(c) *Op. cit.*, p. 39, line 45; plus p. 40, line 49; plus p. 43, line 49.
(d) Based on new classification.
(e) Based on old classification.
(f) 13th Census, Vol. IV, p. 55, division total; less photographers, p. 56; less miners and fishermen (Table 2, Columns A and E); plus electricians (14th Census, Vol. IV, p. 36, line 30).
(g) After transferring all possible items in an endeavor to make the old classification for 1910 equal the new, a deficiency of 766,000 (7.77%) exists, which apparently consists of laborers actually working in these pursuits but reported by the Census as laborers not specified, in the domestic and personal service division.
(h) Column A plus Column B.
(i) 13th Census, Vol. IV, p. 54, division total.
(j) Similar to Note g, a deficiency of 5.04%.
(k) Column E plus Column F.
(m) 12th Census, Sp. Rpt. Occupations, p. LI, division total; plus electricians, p. L (in 1900 only); less photographers, p. LII; less miners and fishermen (Table 2, Columns A and E).
(n) Estimated by applying the per cents mentioned in Notes g and j to Columns A and E respectively, and multiplying these results by 2.82 in 1900, 1.93 in 1890, 2.66 in 1880, and 1.29 in 1870, in order to allocate the balance of the laborers not specified remaining after the transfers previously made to domestic and personal service (Table 3, Note a), and to agriculture (Table 4, Note m).
(o) 12th Census, Sp. Rpt. Occupations, p. L, division total.
(p) Column A has been increased by 30,000 and Column E by 20,000 to allow for faulty enumeration in the Southern States. (See Table 4, Note o.)
(q) Of the persons engaged in manufacturing and mechanical pursuits and in trade and transportation in 1870, 67.1% were in the former and 32.9% in the latter. In 1840, the per cents were 79.3 and 20.7 and in 1820 they were 82.9 and 17.1 respectively. Interpolating on straight lines gives for manufacturing and mechanical pursuits 71.2% in 1860, 75.2% in 1850, and 81.1% in 1830, and for trade and transportation 28.8% in 1860, 24.8% in 1850, and 18.9% in 1830. These per cents are applied to the total number of persons in these two groups, which is calculated by subtracting those listed in other occupations (Tables 2, 3 and 4) from Table 4, Column C.
(r) 6th Census, p. 475, manufacturing and trades.
(s) *Ibid.*, commerce, navigation of the ocean, canals, lakes and rivers.
(t) 4th Census, summary table.

REVISION OF THE WEEKLY INDEX NUMBER

By Irving Fisher

The number of commodities in 1926 is 202, the list being the same as for 1925 except that "Oil: Soya Bean Spot" and "Lumber: F. A. S., Pl. Red Gum" have been dropped. The base period is the last week in December, 1925, and the base number, 159.45 (that being the index number for that week in the 1925 series).

TABLE I

WEEKLY INDEX NUMBERS

(Continuation of Table I, this Journal, September, 1925, p. 410)

For week ending Friday	Index number of prices in per cent of pre-war, i. e., 1913=100 per cent	Purchasing power of the dollar in pre-war cents, i. e., 1913=100 cents
1925 (*Cont.*)		
July 31	160.8	62.2
August 7	161.4	61.9
August 14	160.6	62.3
August 21	158.7	63.0
August 28	157.3	63.6
September 4	158.4	63.1
September 11	158.1	63.3
September 18	159.0	62.9
September 25	158.3	63.2
October 2	157.0	63.7
October 9	156.1	64.1
October 16	156.5	63.9
October 23	156.1	64.1
October 30	157.9	63.3
November 6	158.6	63.0
November 13	158.7	63.0
November 20	159.0	62.9
November 27	160.0	62.5
December 4	161.3	62.0
December 11	159.1	62.8
December 18	157.6	63.5
December 25	157.9	63.3
January 1	159.3	62.8
1926		
January 8	159.3	62.8
January 15	158.7	63.0
January 22	159.2	62.8
January 29	159.7	62.6
February 5	159.6	62.7
February 12	158.6	63.0
February 19	156.6	63.8
February 26	155.5	64.3
March 5	154.0	65.0
March 12	153.3	65.2
March 19	153.6	65.1
March 26	151.7	65.9
April 2	151.8	65.9
April 9	150.7	66.3
April 16	151.6	66.0
April 23	151.2	66.2
April 30	150.2	66.6
May 7	151.2	66.2
May 14	152.3	65.7
May 21	152.9	65.4
May 28	152.6	65.6
June 4	151.8	65.9
June 11	153.6	65.1
June 18	152.3	65.7
June 25	152.0	65.8
July 2	150.2	66.6

TABLE II

MONTHLY AVERAGES

(Continuation of Table II, this JOURNAL, September, 1925, p. 411)

	Index Number 1913=100	Purchasing power 1913=100
1925 (*Cont.*)		
July	160.3	62.4
August	159.5	62.7
September	158.2	63.2
October	156.7	63.8
November	159.1	62.9
December	159.0	62.9
1926		
January	159.2	62.8
February	157.6	63.5
March	152.9	65.4
April	150.9	66.3
May	152.2	65.7
June	152.0	65.8

TABLE III

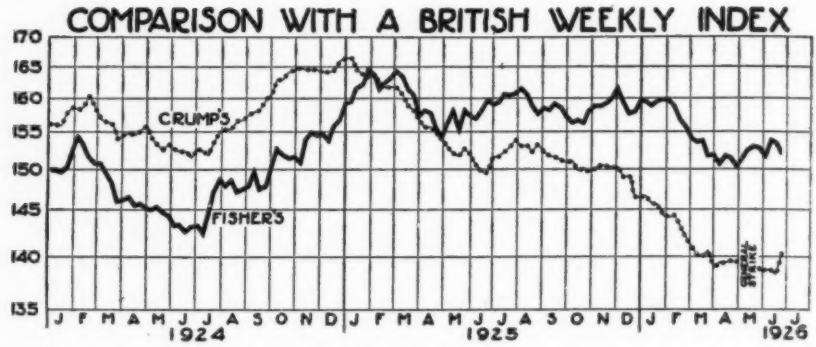
QUARTERLY AVERAGES

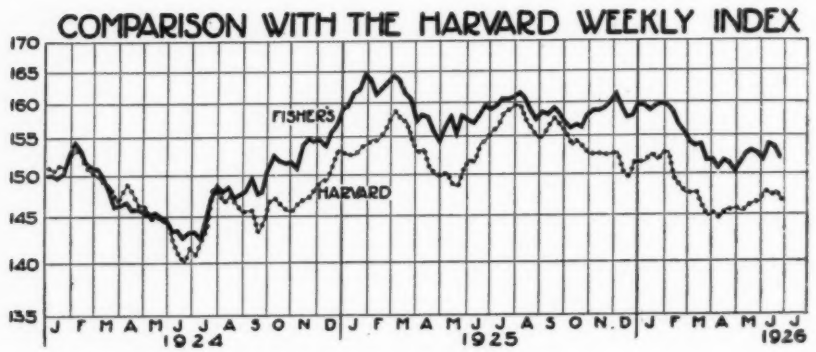
	Index Number 1913=100	Purchasing Power 1913=100
1923		
1st Quarter	161.3	62.0
2nd Quarter	162.6	61.5
3rd Quarter	154.2	64.9
4th Quarter	152.9	65.4
1924		
1st Quarter	150.5	66.4
2nd Quarter	144.7	69.1
3rd Quarter	147.3	67.9
4th Quarter	153.9	65.0
1925		
1st Quarter	161.9	61.8
2nd Quarter	157.2	63.6
3rd Quarter	159.2	62.8
4th Quarter	158.3	63.2
1926		
1st Quarter	156.3	64.0
2nd Quarter	151.7	65.9

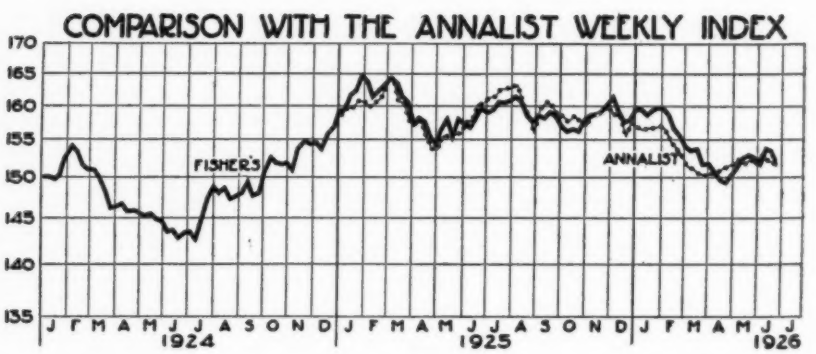
TABLE IV

YEARLY AVERAGES

	Index Number 1913=100	Purchasing Power 1913=100
1923	157.7	63.4
1924	149.3	67.0
1925	159.2	62.8


COMPARISON WITH A BRITISH WEEKLY INDEX
170
165
160
155
150
145
140
135
CRUMP'S
FISHER'S
SPECIAL STUDY
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J
1924
1925
1926


COMPARISON WITH THE HARVARD WEEKLY INDEX
170
165
160
155
150
145
140
135
FISHER'S
HARVARD
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J
1924
1925
1926


COMPARISON WITH THE ANNALIST WEEKLY INDEX
170
165
160
155
150
145
140
135
FISHER'S
ANNALIST
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J
1924
1925
1926

A CRITICISM OF PROFESSOR CHADDOCK'S "PRINCIPLES AND METHODS OF STATISTICS"

To the Editor of the Journal of the American Statistical Association:

Sir:

In Professor Dr. R. E. Chaddock's article on *The Function of Statistics in Undergraduate Training* in the April number of the JOURNAL the following passage occurs:

> The use of the inductive method requires rigorous training in scientific procedures. We are permitting the undergraduate in the social sciences to become lost in the wilderness of concrete data and badly supported general statements without showing him how to find his way about. He has not acquired the wariness and wisdom that scientific training gives, and he is lured constantly on false trails, due to the bias of his hopes and enthusiasms. We have an obligation of first importance to remedy this situation. The elements of statistical methods given as a fundamental discipline to undergraduates will furnish them with the broad foundations for scientific work and will cultivate their critical judgment of facts and explanations offered by others.

All this sounds very nice, and will no doubt appeal to the "Gelehrten" among our various statistical faculties. But a closer examination of the recent deluge of elementary statistical texts which have appeared on the American book market during the past two years oftentimes shows that the "Gelehrten" have done next to nothing in imparting "the wariness and wisdom that scientific training gives" and to which Professor Chaddock so feelingly refers.

Several years ago, while preparing the manuscript for an article on probabilities and generalized frequency functions, the present writer had occasion to make some studies on heights and weights of army recruits and military conscripts of various nations and races. The quantitative analysis was based on expansions in Gram's series. A part of the final numerical results derived from these expansions were later on published in tabular form in Dr. F. L. Hoffman's brochure on *Army Anthropometry and Medical Rejection Statistics*. (Prudential Press, 1918.)

Since some of these results have been subject to what I consider to be an erroneous and misleading interpretation in Professor Chaddock's text on *Principles and Methods of Statistics* it seems but fair to both my old chief and friend, Dr. Hoffman, and to myself that I should be afforded an opportunity to point out misleading interpretations of this sort, especially when coming from an alleged authoritative and learned source as that of one of the "Gelehrten" of the statistical faculty of Columbia University.

I am particularly referring to two sets of numerical tables relating to American army recruits and Japanese conscripts to which Dr. Chaddock has made repeated references in his text and which are taken from a larger table appearing on page 33 of Dr. Hoffman's brochure. In order to show the real and underlying facts of these data I can do no better than to give a few quotations from Dr. Hoffman's pamphlet.

After pointing out that the mean or average height of various races is often useless for practical purposes, Dr. Hoffman continues as follows:

> The typical frequency distribution of stature is, however, of decidedly greater scientific value; and as useful contributions to a better understanding of this important aspect of army recruiting a series of tables and frequency curves has been prepared for European conscripts and American recruits. A few typical frequency curves of stature for the purpose of convenient illustration have been worked out for me by Arne Fisher. . . . Once having computed the various statistical parameters represented by the means and dispersion, the skewness and excess the frequency distribution is easily ascertained for any value of the varying attributes and reduced to a common standard of measure (in this instance to English inches). . . . Computing the distribution from the equations of the frequency curves, Mr. Fisher presents the following comparative results on the basis of 1000 standard measurements; progressing by one-inch intervals for the six countries for which data were available.

The above statement shows conclusively that the comparative results referred to by Dr. Hoffman are *probable* or *presumptive* frequency distributions and *not actual observations*. This fact is further emphasized in the table itself where the frequencies are given with one decimal, which to even the most simple undergraduate student should be conclusive evidence that the values were those of theoretically determined distributions and *not* the actual observations themselves. Moreover, any person with the least bit of that critical sense, which Dr. Chaddock seems to value so highly, could at least from the Japanese data draw the conclusion that the numbers were derived from mathematical equations. Japan uses namely the metric system, and the actual observations would therefore be expressed in centimeters and not in intervals of English inches.

As a matter of fact the only collection for which I had observations available in sufficiently small intervals so as to allow a direct computation of the parameters was that relating to nearly 19,000 Danish conscripts for the year 1916. For the other materials the data as originally submitted to me were given in relatively large as well as unequal intervals so that in order to get a reliable idea of the probable frequency distributions it became necessary to resort to an expansion in a Gram series by means of least square methods in conjunction with numerically determined weights for the unequal intervals of the observations. As a common generating function I used the Laplacean function of the Danish data and determined the coefficients c_0, c_1, c_2, c_3 and c_4 of this generator and its successive derivatives by means of the method outlined by Gram. Once these constants were found it was a simple matter to make c_1 and c_2 vanish by placing the origin of the coördinate system at c_1, or the mean, and by letting the unit of the coördinate system equal c_2. This gave me the parameters from which the table in Dr. Hoffman's pamphlet was computed.[1]

So much for the prior work done by Dr. Hoffman and myself on this subject. Let us now see how these results have fared in the hands of Dr. Chaddock whose

[1] For those interested in the actual crude statistics which were at my disposal in the case of the U. S. A. material the following table pertaining to 304,113 white and 18,002 colored recruits is presented:

Inches	63.5 or less	63.5–66.5	66.5–68.5	68.5–71.5	71.5 and over	Total
Obs. No.	3,527	107,553	109,643	86,361	15,031	322,115

text according to his own words is "aimed to appeal to the logical faculties of the reader and to foster a healthy scepticism towards the results of quantitative investigation."

From the above mentioned table in Dr. Hoffman's brochure Professor Chaddock has taken the presumptive numerical values, suppressed the significant decimal points, and considered these purely probable or presumptive values as the *actual* observations on the respective heights of 10,000 American recruits and 10,000 Japanese conscripts.

The heading of table 46 on page 224 of Professor Chaddock's text states specifically that he considers the presumptive error curves as being *actual distributions*. This in spite of the fact that a reader with even the most rudimentary sort of statistical knowledge would pause to wonder how it was possible to obtain 10,000 measurements on Japanese soldiers expressed in intervals of one English inch when we know that the official code of measure in Japan is that of the metric system.

But the fun becomes even better as Professor Chaddock proceeds with his analysis of what he considers actual measurements, but which really represents theoretically determined presumptive values of the frequency distributions. Without the batting of an eyelash he proceeds to "fit" these theoretically determined Gram series to normal curves. A rather futile procedure when we bear in mind that the normal error law is but the first term of the general Gram series. Next he plots the presumptive values of the Gram series as frequency polygons, still believing that he is dealing with actual measurements, and superimposes upon these polygons his normal frequency curves. And now for the final blurt which I shall quote in full from Chaddock's book and wherein he says: "Comparison of the Japanese and American polygon and curves show that the actual distributions for the latter do not conform quite so closely to the bell shaped curve, but the difference is *not so great as to indicate a different law of distribution of heights.*" (Italics are mine.)

I can hardly think of a more misleading and erroneous statement than this one. There exists a decided and *real* skewness in the American data. Its numerical value in the Gram series is -0.057 with a probable error of $0.2753 \div \sqrt{322115} = 0.00048$ where 322115 is the number of American recruits actually measured and which entered into my determination of the generalized Gram series. If we employ Pearson's definition of skewness we find it to be equal to $+0.171$ since the Pearsonian unit is about minus one third times the *directional* unit employed by Gram. The mode of the American curve is approximately $67.52 + 3c_3\sigma = 67.14$ inches, where $c_3 = -0.0570$ is the measure of skewness according to Gram's method and for which the numerical value is given on page 31 of Hoffman's brochure. This is practically a whole centimeter less than the mean and mode in a normal distribution and constitutes quite an important deviation in anthropological measurements involving over 300,000 individual observations.

In the way of contrast to the Chaddockian analysis I beg to offer the following lucid remarks by Dr. Hoffman on the same subject:

> More precise statistical analysis readily disproves the common error that all statistical frequency curves are true symmetrical curves. This point of view was first advanced by the German mathematician, Gauss, and unfortunately accepted among statisticians of modern times. As a matter of fact, however, the symmetrical distribution is the exception rather than the rule, and the correct ascertainment of frequency distribution requires in addition to the mean and the dispersion the computation of at least two additional parameters, the skewness and the excess.

Statements like this one by Dr. Hoffman shows that the worth while statistical analysis is being done by less learned but far more intelligent fellows—men interested in their facts and their auxiliary mathematical methods more than in the platitudes of the sociological statisticians. That is perhaps why these non-academic fellows reach more valuable results than the majority of the brotherhood of our "Gelehrten."

ARNE FISHER

To the Editor of the Journal of the American Statistical Association:

Sir:

I wish to thank you for allowing me to see the criticism of my book, *Principles and Methods of Statistics*, submitted to this JOURNAL by Mr. Arne Fisher. For the information of those who are interested I would like to state that arrangement has been made with the publishers, in the second printing which is just going to press, to include in the footnote to Table 46, page 224, the following:

> The term *actual* is used in contrast with *ideal*. In the source cited the data, here designated as actual distributions, were obtained through the use of mathematical processes applied by Arne Fisher to the original observations, transformed from metric units to inches, and arranged in intervals of one inch.

At the time of original printing this was consciously omitted since it was clear in the source cited, and our purpose was merely to present the *technique* of fitting the Gaussian curve.

Faithfully yours,

ROBERT E. CHADDOCK

A BASIS OF FORECASTING NATIONAL INCOME

It is entirely logical that the national income should fluctuate in proportion to bank deposits. A comparison of the estimates of national income computed by Dr. Willford I. King and Oswald W. Knauth of the National Bureau of Economic Research with the deposits of "all banks" of the United States seems to demonstrate similarity of fluctuation. Income averaged 95 per cent over deposits from 1910 to 1921 inclusive, and with relatively slight deviation from this average. During the four pre-war years when commodity prices were stable and both deposits and income fluctuated little, the average ratio was 1.95. In 1914, when income suffered from credit restriction after August 1, the ratio fell to 1.75. In 1915, a year of gradual revival, the ratio rose to 1.87. During 1916–19, years of unrestricted inflation, the ratio averaged 2.05. From 1.98 in 1920 the ratio dropped to 1.78 in 1921, a year when credit was "frozen" and the turnover of deposits was slow.

Year	Bank Deposits * (Billions)	National Income † (Billions)	Ratio, Income to Deposits
1910	$15.3	$31.1	2.03
1911	16.0	31.2	1.95
1912	17.1	32.4	1.90
1913	17.5	33.3	1.91
1914	18.6	32.5	1.75
1915	19.2	35.9	1.87
1916	22.8	45.5	2.00
1917	26.4	53.9	2.04
1918	28.8	61.7	2.14
1919	33.2	67.3	2.03
1920	37.5	74.2	1.98
1921	35.2	62.7	1.78
			1.95 (12-year average)

* From the Annual Reports of the Secretary of the Treasury. The series includes individual deposits and United States deposits, as of June 30, for all banks in the United States except the 12 Federal Reserve Banks.

† From *Income in the United States*, National Bureau of Economic Research, Vol. 1, p. 13 (data for 1910–1918); Vol. 7, p. 32 (data for 1919–1921).

Since 1921 deposits have increased and are substantially above 1920, the actual figures rising from 37.5 billion dollars in 1920 to 46.9 billion in 1925. This is a gain of 25 per cent. A corresponding increase of 25 per cent in the national income would mean that the actual figures had risen from 74.2 billion dollars in 1920 to 92.8 billion in 1925. Such a large national income for 1925 will be questioned by those who give attention primarily to the current income from crops, railroad revenues, sales of iron and steel, copper, etc. The total payrolls and incomes in transportation, agriculture, iron and steel and some other industries have not increased anything like 25 per cent since 1920. On the other hand, there has been a large increase in new building, petroleum output, automobile production and a number of other sources of income. Including all the elements, it is possible that the national income for 1925 approximated 92.8 billion dollars. A comparison with actual statistics of bank deposits for the past six years supports this estimate. The figures in the table, covering both years of stable commodity prices before the war and years of phenomenal wartime inflation of prices, show that the national income has averaged about 95 per cent above deposits. Adding this percentage to deposits for the years 1922 to 1925 inclusive, we have the following estimates of national income:

INCOME ESTIMATED, CONSTANT RATIO ASSUMED

Year	Bank Deposits * (Billions)	National Income (Billions)	Ratio, Deposits to Income
1922	$37.3	$72.7	1.95
1923	40.3	78.6	1.95
1924	43.1	84.0	1.95
1925	46.9	91.5	1.95

* Annual Report of the Secretary of the Treasury, 1925, p. 369; includes "Individual deposits" and "United States deposits" as of June 30.

This table, of course, needs revision. A uniform ratio of 1.95 should not be used in all years. Comparing years of trade activity and trade depression before the war we find a ratio of 2.03 in 1910 and 1.75 in 1914. Similarly, each year dur-

ing 1922–25 requires special study as to trade activity, money market conditions and price movements. Arbitrarily we shall employ 1.87 for 1922 when recovery was slow, 1.95 for 1923 and 1924, 1.98 for 1925, 1.95 for 1926, and 1.92 for 1927. We shall also assume an increase of 2.8 billion for deposits in 1926, and a gain of 2.0 billion for 1927. We then have the following estimates for 1922–25 and a forecast of 96.9 billion dollars for the national income in 1926 and 99.3 billion for 1927.

INCOME ESTIMATED, VARIABLE RATIO ASSUMED

Year	Bank Deposits (Billions)	National Income (Billions)	Ratio of Income to Deposits
1922	$37.3	$69.8	1.87
1923	40.3	78.6	1.95
1924	43.1	84.0	1.95
1925	46.9	92.9	1.98
1926	(est'd) 49.7	96.9	1.95
1927	(est'd) 51.7	99.3	1.92

In addition to the comparison of deposits with income, it is interesting to note that bank clearings "outside of New York" during 1910–1925 have averaged 4.6 times deposits. In years of credit restriction, such as 1914 and 1921, clearings have fallen to a 4 to 1 basis, while in 1919, 1923 and 1925, years of speculative activity, a 5 to 1 ratio has been approximated. The following table shows the three series, giving both actual figures and relatives (income for 1922–25 estimated).

Year	Deposits		Outside Bank Clearings		National Income	
	Actual	Relatives	Actual *	Relatives	Actual	Relatives
	(Billions)		(Billions)		(Billions)	
1910	$15.3	100	$66.8	100	$31.1	100
1911	16.0	105	67.9	102	31.2	100
1912	17.1	112	73.2	110	32.4	104
1913	17.5	115	75.2	112	33.3	107
1914	18.6	122	72.2	108	32.5	105
1915	19.2	126	77.3	116	35.9	116
1916	22.8	149	102.3	153	45.5	146
1917	26.4	173	129.5	194	53.9	173
1918	28.8	188	153.8	230	61.7	199
1919	33.2	217	182.0	274	67.3	216
1920	37.5	245	208.3	312	74.2	239
1921	35.2	231	143.9	215	62.7	202
1922	37.3	244	157.8	237	69.8	224
1923	40.3	264	199.5	299	78.6	253
1924	43.1	282	205.2	307	84.0	270
1925	46.9	307	226.0	338	92.9	299

* As compiled by the *New York Commercial and Financial Chronicle.*

On the gold base provided by the current reserves of the Federal Reserve Banks, deposits can be increased with facility for several years at an average rate of 2 billion dollars or more per annum. Ten years hence, however, the credit superstructure should begin to show maturity relative to gold reserves under existing legal reserve requirements. Deposits will then probably increase at a

lower annual rate and the per capita money income of the country, therefore, will probably grow less rapidly. Real income, however, should then improve more rapidly than at present. Inventions and improvements in labor-saving machinery contribute more to real income than does credit inflation, and they multiply most rapidly when speculation is dormant and credit expands no more than commercial growth requires.

WARREN F. HICKERNELL

New York University

A SCALE FOR PLOTTING LOGARITHMIC CURVES

Charts employing a logarithmic scale, for one or both dimensions, are now in very common use. For plotting curves on such charts, logarithmically ruled paper is widely used and is often convenient. Such paper is, however, subject to inherent limitations of inflexibility as well as to other defects which make it unsatisfactory in many cases. Plotting logarithms of the data with appropriate grid lines is accurate and permits ready adjustment of proportions but is laborious. Many statisticians, therefore, have prepared logarithmic scales of paper or cardboard or have resorted to still other devices to facilitate direct plotting of data on the logarithmic scale.

Apparently many are not yet aware that there is commercially available a convenient combination of logarithmic scales which serves most purposes better than any of home manufacture. Several years ago, at the writer's suggestion, Keuffel & Esser Co. made up a triangular rule of their "Paragon" series on which were set off six different logarithmic scales containing the graduations employed on the company's different slide rules. The most open of these scales gives a single deck about 20 inches long; the least open a single deck about 1⅝ inches long. With this range of selection it is possible to plot with ease and precision almost any logarithmic chart needed for working purposes or for publication.

The Food Research Institute has used this rule for several years with entire satisfaction and believes that other organizations and individuals may find it a valuable addition to their equipment.

J. S. DAVIS

Food Research Institute, Stanford University

PROGRESS OF WORK IN THE CENSUS BUREAU

ANNUAL STATISTICS OF INSTITUTIONS

Various organizations interested in securing more adequate statistics of criminals and of the mentally diseased and the mentally defective have been urging for some time that the Federal Government collect annually statistics of the institutions for these classes somewhat like the figures which have heretofore been collected by the Bureau of the Census once in ten years. A conference held at the Bureau of the Census on April 28 was attended by representatives of these organizations as follows: Sanford Bates, President of the American Prison Associ-

ation and Commissioner of Correction of Massachusetts; E. R. Cass, General Secretary of the American and New York State Prison Associations; Frederic B. Crossley and Sam B. Warner, representing the American Institute of Criminal Law and Criminology; Emil Frankel, Statistician of the Pennsylvania State Department of Welfare; Hastings H. Hart, of the Russell Sage Foundation; Horatio M. Pollock, Statistician of the New York State Hospital Commission and representing the American Psychiatric Association and the American Statistical Association; George K. Pratt and Frankwood E. Williams of the National Committee for Mental Hygiene.

As a result of this conference, it was agreed that the Bureau of the Census should undertake the collection of annual statistics, limiting the inquiry at first to state penal institutions, state hospitals for the insane, and state institutions for the feeble-minded and epileptic. It is proposed to obtain the information direct from the institutions, except in those states where it may prove advantageous to have the information furnished by the State Department of Welfare or other state office having supervision over the various institutions. Tentative schedules have been worked out and will shortly be sent to the institutions and to the various state offices. It may be necessary to make the first complete report relate to the year 1927, since some of the institutions may need to modify their system of records in order to provide all of the items that will be called for. It is expected, however, that a report of some kind will be issued for the year 1926, even though this may be less extensive in scope than the reports for subsequent years.

The information to be obtained from each institution will comprise the following:

1. A statement in balance-sheet form showing the number of inmates at the beginning of the year, the number of admissions, the number of departures, and the number in the institution at the end of the year.

2. The number of admissions classified in some detail as to sex, race, nativity, and age; as to offense, sentence, and previous convictions, for prisoners; or as to mental classification for the insane and mentally defective.

3. The number of departures with a simpler classification.

So far as classification is concerned, the major part of the emphasis will be placed on admissions, though information will be obtained as to the conditions under which persons leave the institutions. For example, persons leaving a penal institution will be classified as leaving at expiration of sentence, by parole, by pardon, etc., and patients leaving hospitals for the insane will be classified according to condition on leaving, as recovered, improved, unimproved, etc.

MISCELLANEOUS NOTES

United States Bureau of Labor Statistics.—Field work is well under way in union scales of wages; wages and conditions of labor in the men's clothing industry; in the textile industries—cotton and wool.

An intensive study of the apprenticeship system in the building trades is under way. This study will later be extended to other industries.

The labor productivity indexes referred to in the last issue of the JOURNAL are being pushed as rapidly as possible. This work is under the direction of Professor Ewan Clague, formerly of the University of Wisconsin. The July *Monthly Labor Review* contains the first installment of Mr. Clague's studies. The methods he there describes are now being applied to a number of other industries.

Mr. J. C. Bowen has returned from his trip to Europe which has proved very successful both as to the pottery industry and to the iron and steel industry, the two studies which he had in hand.

Home Study Courses in Vital Statistics for Public Health Workers, Columbia University.—Under the auspices of the Committee on Educational Standards of the Section on Vital Statistics, American Public Health Association, a home-study course on descriptive vital statistics is being planned in cooperation with Professors Robert E. Chaddock and Charles H. Hall of Columbia University. The outline of the course is being made by Dr. W. J. V. Deacon and Mr. Edwin W. Kopf. It is hoped that full details regarding the course may be placed before the meeting of the American Public Health Association at Buffalo, New York, October 11–14, 1926. For several years the Section on Vital Statistics of that Association has been considering ways and means for providing instruction in vital statistics for health officials, registrar's assistants and for persons in public and private service whose work deals more or less intimately with population and vital statistics. The course will be based upon the texts of Newsholme, Whipple, Pearl and Chaddock. Incidental readings will be assigned in the writings of Farr and his successors in England; of Edward Jarvis, John Shaw Billings, Henry Brooks Baker, Ezra M. Hunt, Cressy L. Wilbur, John W. Trask, L. I. Dublin, Walter F. Willcox and Frederick L. Hoffman in the United States; George H. Knibbs of Australia; Harold Westergaard of Denmark; Georg von Mayr, Eugen Wurzburger and E. von Roesle of Germany; Dunlop of Scotland; and other authorities whose works are available to American students in the libraries of the larger cities. Special emphasis will be placed upon the introductory and explanatory text in the United States Census volumes on population and vital statistics, 1850 to 1920. For students contemplating advanced work in the analysis of vital statistics, the home-study courses in mathematics now being offered by Professors Brett and Northcott at Columbia University will be recommended. These courses cover the essential mathematical discipline which is prerequisite to the study of statistical analysis. The home-study courses in related public health subjects now being outlined by Dr. Emerson will also be of service to these students. The report of the Committee to the Vital Statistics Section of the American Public Health Association in October will be of interest to teachers of statistics generally.

Meeting of the American Public Health Association.—A meeting of the American Public Health Association will be held at Buffalo, New York, October 11–14, 1926. The various committees will make their reports, and among the subjects discussed in the vital statistics section will be: Death-rates among physicians; some aspects of infant mortality as shown in Canadian Vital Statistics; need of trained statisticians in the hospital field; development of statistics of marriage and divorce in New York State; need for improved vital statistics; forecasting of communicable diseases; one dozen years of vital statistics in North Carolina; constitutional, statutory and administrative aspects of vital statistics practice in the United States; and the use of records from public health nursing groups. Information regarding the forthcoming meeting may be obtained from Mr. Homer N. Calver, Secretary, American Public Health Association, 370 Seventh Avenue, New York City.

Pennsylvania Department of Labor and Industry.—During the year 1926 the Bureau of Statistics of the Pennsylvania Department of Labor and Industry is working for the definite accomplishment of the following purposes:

1. Centralization of the assembly and distribution of all statistics related to labor and industry.
2. Revision and standardization of industries classification.
3. Complete revision of accident cause code so that it will list types of machines more in detail and consequently permit a more accurate determination of actual hazards involved in connection with operation of specific types of machines.
4. Correlation of current employment and accident data.
5. Extension of scope of collection of employment and wage data so as to include a fair sample of the construction industry.
6. Collection of data of labor turnover.
7. Compilation of accident experience for individual concerns and computation annually of accident frequency and severity rates based on man hour exposure. Possibly the experience of a representative group of firms in each industrial class will be published with the identity of individual concerns concealed.
8. Studies of the industrial accident problem as it affects women and children.
9. Practical and comprehensive use of statistical material now available rather than expansion of collective functions.

Employment Statistics.—The report setting forth the results of the work of the Committee on Governmental Labor Statistics of the American Statistical Association, of which Mary Van Kleeck of the Russell Sage Foundation is Chairman, is being published as *Employment Statistics for the United States.* It contains a plan to forestall periods of extensive unemployment, such as often follow waves of unusual prosperity. Its purpose is to lessen unemployment by providing facts needed to understand and control fluctuations in the production of goods and obstructions to their sale. It is expected that the book will be reviewed in an early issue of the JOURNAL.

From July 1, a *Monthly Review of Current Business Statistics* will be issued in print for Canada by the Dominion Bureau of Statistics, at a price of $1.00 per year. Since the beginning of the present year this publication has been put out in mimeographed form, but the demand for it has become so general as to warrant printing.

Under the leadership of Professor A. H. Mowbray a new chapter of the American Statistical Association was organized at San Francisco on June 9, 1926.

A Canadian Statistical Association, which aims to include membership throughout the Dominion, is in process of organization. A local statistical association was organized some months ago in Montreal and it is through the initiative of this body that the larger field will now be explored. The formation of a local statistical organization at Toronto is also under discussion.

Monthly reviews of motor car production and of boot and shoe production in Canada have lately been instituted by the Dominion Bureau of Statistics.

Dr. Burton C. Camp, Professor of Mathematics at Wesleyan University, has been a non-resident lecturer in Mathematical Statistics in the University of Michigan Summer Session. He gave an elementary and an advanced course both of which had an unusually large attendance. Dr. Camp recently completed a year of study under Karl Pearson.

Professor S. D. Wicksell, of the University of Lund at Lund, Sweden, has been appointed special lecturer at the University of Michigan to be in residence the entire second semester beginning in February, 1927. Professor Wicksell will give an advanced course in Frequency and Correlation Methods, a course in Stellar Statistics, and a seminar on research in various lines of application of Mathematical Statistics.

Since the return of Professor Corrado Gini from America, where he served as a member of the War Debts Delegation, he has been transferred to the Professorship of Economic Statistics and Economic Politics at the University of Rome.

Professor Ray B. Westerfield, of Yale University, has been commissioned with the task of organizing the statistical work of the Federal Treasury Department.

Mr. Joseph E. Pogue has been appointed Vice-Chairman of the Petroleum Division of the American Institute of Mining and Metallurgical Engineers, in charge of Petroleum Economics.

RECENTLY ELECTED HONORARY MEMBERS OF THE AMERICAN STATISTICAL ASSOCIATION

Sir Josiah Stamp, K. B. C., Tantallon, Park Hill Road, Shortlands, Kent, England
Professor Emanuel Czuber, Technical High School, Vienna, Austria
Dr. H. W. Methorst, Director, Central Bureau of Statistics, The Hague, Holland
Dr. T. H. C. Stevenson, Medical Superintendent, General Register Office, Somerset House, W. C., London, England
Dr. E. Roesle, Editor, *Archiv für Soziale Hygeine und Demographie*, Cimbernstrasse 6, Berlin, Germany
Dr. Friedrich Weyr, President of the State Statistical Office, Prague, Czechoslovakia

MEMBERS ADDED SINCE JUNE, 1926

Ackerson, Dr. Luton, Psychopathic Hospital, Iowa City, Iowa
Aull, Robert P., Willys-Overland Inc., Detroit, Mich.
Beere, Thekla J., Irish Free State Statistical Bureau, In care Laura Spelman Rockefeller Memorial, 61 Broadway, New York, N. Y.
Bennett, James B., Bureau of Efficiency, Washington, D. C.
Bittner, Gustave E., Domestic Commerce Division, Bureau of Foreign and Domestic Commerce, Washington, D. C.
Burmeister, Gustave, Bureau of Agricultural Economics, Cheyenne, Wyo.
Burr, Dr. Susan S., Economic Research in Public Finance, Washington, D. C.
Burrows, Robert N., Department of Commerce, Bureau of Fisheries, Washington, D. C.
Carson, William Joseph, Federal Reserve Board, Washington, D. C.
Cooper, P. F., General Motors Export Company, 224 West 57 Street, New York, N. Y.
Cooper, Walter W., 436 Crescent Street, N. E., Grand Rapids, Mich.
Cox, Professor Garfield V., School of Business Administration, University of Chicago, Chicago, Ill.
Dayton, Willard L., Hudson Motor Car Company, Planning Department, Detroit, Mich.
Didriksen, C. N., Automobile Finance, Detroit, Mich.
Dodge, Harold F., Bell Telephone Laboratories, Inc., New York, N. Y.
Eastwood, R. Parker, 521 West 123 Street, New York, N. Y.
Edie, Professor Lionel D., Director, Bureau of Business Research, 620 East First Street, Bloomington, Ind.
Edwards, Mary Stockton, American Social Hygiene Association, New York, N. Y.
Ellis, Wayne P., Ellis Coal Bureau, Minneapolis, Minn.
Fourcade, Germain A., Guaymas, Sonora, Mexico
Gibbs, J. Barnard, State Department of Agriculture, Charleston, W. Va.
Goodwin, Edward S., Analysis Insurance, Bank and Trust Stocks, 962 Main Street, East Hartford, Conn.
Gourvitch, Alexander, 3224 Grand Concourse, New York, N. Y.
Handler, Edith H., 313 Sterling Hall, University of Wisconsin, Madison, Wis.
Harmon, Julia, 411 West 116 Street, New York, N. Y.
Heard, Herman G., U. S. Department of Agriculture, Houston, Texas
Horner, S. L., School of Business Administration, University of Michigan, Ann Arbor, Mich.
Hussey, Miriam, Industrial Research Department, University of Pennsylvania, Philadelphia, Pa.
Johnson, Miss India G., S. J. Stern and Co., New York, N. Y.
Kaufmann, H. J., Assistant Director, Board of Education, In charge of Statistics and Publications, Detroit, Mich.
Koehler, Richard T., School of Business, Columbia University, N. Y.
Leuschner, Richard D., Chace-Leuschner, Inc., Real Estate, Ocala, Fla.
Lyman, Margaret Helen, New York Tuberculosis and Health Association, 244 Madison Avenue, New York, N. Y.
MacMillin, Frederick N., Municipal Reference Library, Milwaukee, Wis.
Merry, Dr. Glenn Newton, Market Research, Brown Co., Portland, Me.
Mitchell, Harcourt, General Outdoor Advertising Co., 17 Park Avenue, New York, N. Y.

Montgomery, Wm. L., Analysis and Research, Federal Reserve Bank of San Francisco, San Francisco, Calif.

Otto, Charles, Jr., Department of Economics and Statistics, Brooklyn Edison Company, Inc., 634 Bergen Avenue, New York, N. Y.

Pryor, Millard H., Keane, Higbie and Co., Investment Bankers, Detroit, Mich.

Richards, Robert G., Manhattan Life Insurance Co., 66 Broadway, New York, N. Y.

Richmond, E. Grace, Sociology Department, Ohio State University, Columbus, Ohio

Rogers, Herbert O., Dennison Manufacturing Co., Framingham, Mass.

Rolph, John G., Private Research, Finance and Industrial Statistics, 35 East 9 Street, New York, N. Y.

Schmidt-Ernsthausen, Dr. Friedrich Maximilian, Department of Agricultural Economics, Cornell University, Ithaca, N. Y.

Smith, Roscoe Israel, Thomson and McKinnon, The Rookery, Chicago, Ill.

Snell, Julia Charlotte, 302 Montgomery Avenue, West Pittston, Pa.

Standish, Alexander, Investment Trust of America, Detroit, Mich.

Staub, Elmer E., Public Accounting, E. E. Staub and Co., Detroit, Mich.

Studt, Ward B., University of Denver, Denver, Colo.

Sun, Tsung Yu, Graduate Student, Columbia University, New York, N. Y.

Swift, Linton B., American Association for Organizing Family Social Work, 130 East 22 Street, New York, N. Y.

Watson, Lucia Norwood, Case Pomeroy and Co., 60 Beaver Street, New York, N. Y.

Weaver, Donald V., 100 North 8 Avenue, Whitestone, N. Y.

White, C. M., State Department of Agriculture, State House, Augusta, Me.

Whittemore, Alice, Statistician, New York Tuberculosis and Health Association, 135 West 16 Street, New York, N. Y.

Wood, Dr. Edith Elmer, Writer and Lecturer on Housing, Cape May, N. J.

Wright, Dr. Sewall, Associate Professor of Zoology, University of Chicago, Chicago, Ill.

Yntema, Theodore O., School of Commerce, University of Chicago, Chicago, Ill.

REVIEWS

Statistical Analysis, by Edmund E. Day. New York: The Macmillan Company. 1925. xxvii, 450 pp.

"Statistics" has well been classified into statistical facts, statistical methods, and statistical theory. American college courses in the subject and the current bumper crop of textbooks deal almost exclusively with statistical methods. For as everyone knows, statistical facts are in themselves of very limited interest; and statistical theory is an esoteric mystery of the Eastern Hemisphere which has made small headway on this continent. And the statistical methods which hold the stage are methods of a very special sort; for the teaching of statistics in North America has been much under the influence of what R. A. Fisher in a recent book calls "the painful misapprehension that statistics is a branch of economics."

The annual drilling of thousands of students in methods without theory is buttressed by the argument that theory is for specialists and that what the ordinary person needs is a good working knowledge of practical methods. The same argument would logically omit trigonometry but give college students generally a course in surveying, and replace psychology in the curriculum by salesmanship.

This idea, the reviewer ventures to hold, is about as far as possible from the truth. Intelligent men and women not specializing in a subject can often learn something about it with profit—else why the liberal college?—but that something is theory, a consideration of the logical relations of the subject to the general body of knowledge and of its parts to one another, rather than detailed technique. College students not specializing in engineering will gain more by a course in electromagnetic theory and atomic structure than by a course in winding armatures and installing doorbells, though the latter kind of knowledge has a quotable market value and the former has none. The electrical engineering student should have both kinds of knowledge. Similarly it is of more value to educated men generally to know the derivation of a probable error formula—a derivation making clear the underlying assumptions and consequent limitations of the result—than to acquire skill in numerical calculation of probable errors even in the most expeditious manner with the best machines. The professional statistician should of course be a master of both theory and technique.

Actually the reason for the neglect of theory is the pervasive ignorance of mathematics which characterizes this country. The European writer, whether British, French, Dutch, German or Scandinavian, can normally assume in developing his theory a fairly intimate acquaintance on the part of his readers with branches of mathematics of which the typical American college student has never heard. In France there are even politicians who, when out of power, make notable contributions to pure mathematics. But there is no use in referring an inquisitive American student to memoirs whose every page is covered with strange

symbols and syllogisms with "now it is obvious that" premises whose obviousness depends on the reader's ability to do calculus mentally.

The writer of an American textbook, no matter how good his intentions, is thus driven to the elimination of all mathematics and consequently of almost all theory. If he is writing a book on statistics he must devote his space to the collection, tabulation and manipulation of data in particular kinds of statistical enterprises. From such texts and the accompanying class work the student derives a contribution to his education comparable to that of the elements of the carpenter trade which have now crowded algebra from the grammar schools.

But since bricks must be made, straw or no straw, and since statistics must be taught, the only thing to do is to make the best of it, and that is what Dean Day has done. Relegating to appendices the collection and tabulation of data, he devotes the bulk of the book to that part of statistical method most closely connected with theory and of most general interest. As he says in the preface:

> The emphasis of the book is upon the analysis, rather than upon the collection and tabulation, of statistical material. It is the author's opinion that the great majority of students of statistics are not directly interested in the technique of collection and tabulation. They are concerned with these phases of statistical method only in so far as they affect the subsequent analysis of data. . . . attention may be profitably focused at the outset on the problems of analysis, and this is the plan of the present volume.

Our hope for a book on statistical theory is encouraged by an interesting and noteworthy paragraph of the preface:

> In more extended form, the title of this book might be given as "The Logic of Statistical Analysis." Several years of experience in teaching statistics have led the author to believe that it is much more important for the beginning student to learn the logical setting of statistical analysis than it is for him to gain facility in any set of technical processes. Some command of technical procedure, of course, is necessary. But, after all, the objectives of statistical work lie in the interpretation of statistical results; and interpretation can never be wisely undertaken save with full recognition of the logical implications and limitations of the processes that have gone before. At the risk, perhaps, of being somewhat meticulous, if not pedantic, the present treatment has been written with the avowed intention of making perfectly clear the *logic* of statistical procedure.

However, the book is a treatise, not on general logical principles or the theory of probable inference, but on the current methods of economic statistics. Meteorology and biometrics supply a few illustrative examples, but in the main the adherence to the economic point of view is pretty close. There is no mention of the applications to psychology, education, preventive medicine, astronomy or statistical mechanics.

As a book on economic statistics it is rather well done. The author's eminence and well-known achievements in the field, coupled with a generally felicitous mode of exposition, add to its value as a text. Recapitulations and a minute classification of subject-matter, though for some readers a nuisance, are good pedagogy. The illustrative examples are genuine, with sources given. The correlation coefficient is introduced in an admirable manner, and the same can be said of the standard deviation. The hundred pages on the analysis of time series give about everything of an elementary nature which is known of the subject and are very good. The chapters on index numbers are also excellent. The exposition of the

bias of arithmetic and harmonic means of relatives is good except that the simple demonstration is omitted. The discussion of the short methods for calculating, by means of an arbitrary origin, the mean, standard deviation and correlation coefficient would be much illuminated by inserting the easy algebraic proofs. Graphic methods do not receive a separate chapter but are considered in connection with the various kinds of analysis of data, an appendix giving detailed rules. There is an extensive use of footnotes for material which most writers would put in the text.

Certain inaccuracies may confuse students, particularly thoughtful students who try to reconcile the contents of the book with what they learn in mathematics courses. "Rate of increase" (Chapter 16), and "term" (of a formula, page 364) are used with meanings in conflict with those established by long mathematical usage. A sentence on page 42 looks like a definition of "function" and the accompanying footnote seems, if it means anything, to mean that every function has an algebraic equation. The second footnote on page 306 loses meaning because of an error in its second sentence.

On page 254 we read that "the angles made with the horizontal axis by different straight lines on logarithmic plot are not to be thought of as proportionate to the [relative] rates of change represented by the different lines." Of course not; but the tangents of these angles are proportional to the instantaneous relative rates of change. The first sentence below the formulae on page 193 is rather mystifying; for if a linear relationship is given how can the correlation be anything but ± 1? The trouble arises from the common practice of confusing population with sample in statistical discussions.

It would be desirable to bring out the fact that frequencies are represented essentially by areas rather than by ordinates. The need of this is particularly great in a book in which distributions with unequal class intervals are discussed at length.

A supply of problems for students to work out would be a valuable feature of an elementary text. As it is the instructor must supply these. But the student's labors will be lightened by the inclusion, among the nine appendices, of tables of logarithms, squares, square roots, reciprocals, and corresponding values of r and ρ.

The book includes no discussion of multiple or partial correlation, curve-fitting or least squares. There is an exposition of how to fit a straight line to data by the usual method, but with no reason given for preferring the result to the many others possible. A concluding chapter on "The Nature of Statistical Results" contains much salutary advice but gives a decidedly inadequate account of the bearing of probability upon statistics.

Perhaps some day we shall come around to the idea that an acquaintance with general statistical theory represented, say, by ability to read *Biometrika*, is of value both to the professional statistician and to the intelligent non-specialist. We shall then have university courses in statistical logic demanding as a prerequisite the necessary mathematics, which involves nothing besides calculus that cannot be mastered in a few days. For specialists in economics or psychology or meteorology there may possibly be advanced courses dealing with applications

to these several subjects, but the materials of such courses will be in a decidedly "boiled down" state. Such a reorganization is brought nearer by the emphasis laid by such a distinguished statistician as Dean Day on the proposition that the part of the method of economic statistics most valuable for college study is a part occupying only a fraction of the energy of particular investigations, the manipulation of the collected data.

HAROLD HOTELLING

Food Research Institute, Stanford University

Causal Factors in Infant Mortality, by Robert Morse Woodbury, Ph.D. United States Department of Labor, Children's Bureau, Publication No. 142. Washington: Government Printing Office. 1925. xi, 245 pp.

This is an excellent study and evaluation of the causal factors in infant mortality. Dr. Woodbury has made a careful analysis of data collected by investigators for the Children's Bureau in eight American cities in selected years between 1911 and 1916. The eight cities include a wide range of industrial activities, and the results found in this study may, therefore, be considered fairly typical of urban conditions throughout the country.

> The data secured in regard to each infant included not only facts available from birth and death certificates—such as sex, whether single or plural birth, number in order of birth, attendant at birth, month of birth, age at death, cause of death, legitimacy, race and country of birth of mother, and age of mother—but information on other points intimately connected with the infant's chances of life, including type of feeding during each month of the first year of life, period of gestation, mother's employment during pregnancy and during the infant's first year of life, father's earnings and occupation, and housing conditions. Facts relating to the mother's whole maternity histories were obtained from which the interval between births could be derived; and in one city, Baltimore, details in regard to the prenatal confinement, and postnatal care received by the mother were sought. (p. 10)

The chief criticism of the data is the doubtful validity of combining different years for different cities, particularly such data as the earnings of the father, which would be vitally affected by changes occurring in wages and employment.

The evaluation of the importance of the various causal factors is made by the use of Westergaard's method of expected deaths (described by Dr. Woodbury in this JOURNAL, Vol. XVIII, pp. 366–376). This method has much to recommend it. It is simple and clear cut, and does not give that illusion of excessive accuracy which is too often created by the use of more complicated statistical analysis. Data of this sort, with inherent inaccuracies and necessary incompleteness, can readily lose their meaning if analyzed by over-refined methods. Dr. Woodbury shows a wholly admirable appreciation of the limitations of his data, and of the impossibility of the complete isolation and final evaluation of the factors involved. His results are, therefore, more trustworthy and have a greater degree of finality than is usual in investigations of this sort.

The primary analysis concerns the immediate or pathological causes of infant deaths and their relation to the factors of age and sex of infant, seasonal condi-

tions, and month of birth. In these eight cities, causes peculiar to early infancy account for one-third of the deaths, gastric and intestinal diseases for three-tenths, and respiratory diseases for one-sixth. The underlying factor of the first of these categories is the condition and care of the mother during confinement. Gastric and intestinal diseases fluctuate with climatic conditions, and are also highly correlated with the type of feeding. Respiratory diseases are also partly a function of climate, and vary greatly among the several national groups. The rates for these categories are, respectively one-and-one-half, six, and four times greater than corresponding rates in New Zealand, which suggests that, with optimum conditions, much infant mortality is preventable.

The few points upon which data were available concerning the physical condition of the mother showed this to be a very important causative factor. "Infants of mothers who died within one year after confinement had a mortality from all causes which was over four times and a mortality from causes peculiar to early infancy which was seven times the corresponding rates for those whose mothers lived for a year or more after confinement" (p. 3). The data are not sufficiently numerous to explain the causes underlying this condition, but a large excess remains even after allowing for an excess of premature births and artificial feeding among this group. There was an excessive mortality of infants born to mothers under 20 and over 40. For all except first births, however, this excess was largely accounted for in young mothers by short interval births and low economic status and in old mothers by relatively great economic pressure and high order of births. The interval between births is also important, the mortality rates being highest for infants born at short intervals after preceding births, and lowest for infants born at intervals of four years or more. This difference could only partly be explained by differences in the age of the mother, order of birth and economic factors. An infant's chance of life was also diminished if the mother became pregnant during its first year of life—largely because of the necessity of the artificial feeding of such infants. Mortality varied also with the type of birth, being excessive for the prematurely born, for twins and triplets, and for infants whose mothers were delivered instrumentally.

An extremely important cause, underlying many others, was the type of feeding. "The mortality among the exclusively artificially fed averaged between three and four times that among the exclusively breast fed." (p. 5)

Differences in racial inborn vigor or vitality are suggested by the great differences existing as among the several national groups, although Woodbury points out the impossibility of isolating these factors from the prejudices and customs of the groups. Thus, mortality rates among the Portuguese were almost four times as great as among the Jews. Two important underlying factors are variations in the degree of artificial feeding and differences in economic status as among the several groups.

Housing congestion was an important causative factor, even after allowing for differences in the national groups and in the earnings of the father.

Of the direct economic factors, the employment of the mother during pregnancy was important. The mortality among infants of mothers employed outside the home was highest, of those gainfully employed in the home next highest, and of

those not gainfully employed lowest. The employment of the mother away from home exerted an influence "irrespective of any correlation with nationality of mother or with father's earnings." (p. 139) The employment of the mother during the first year of the infant's life was found to be an effective cause of mortality only when the employment was outside the home. This was partly explained by a tendency towards early artificial feeding of such infants, and by the overweighting of the group with certain nationalities and with infants whose fathers earned low incomes. Even allowing for all these factors there was an excess "due probably to lack of the care which only the mothers who remained at home could give." (p. 146)

The earnings of the father are closely correlated with infant mortality rates. The chain of causation linking the father's earnings and infant mortality showed an intimate connection between the extent and quality of prenatal care, the gainful employment of the mother during pregnancy, the greater dangers of artificial feeding in the lower income groups (although artificial feeding is less frequent among these groups), the unsatisfactory housing conditions of the lower income groups and the employment of the mother during the first year of the infant's life.

This summary of Dr. Woodbury's results suggests the tremendous complexity of the data, and the close interrelationships of the various causes of infant mortality. His results suggest the theoretically preventable nature of a large part of infant mortality—so much of which is ultimately to be attributed to poverty and ignorance. How far it can be reduced practically is a question for the future. Maternity clinics, education of mothers, birth control, and various forms of social insurance suggest themselves as possible partial solutions.

DOROTHY SWAINE THOMAS

Social Aspects of the Business Cycle, by Dorothy S. Thomas, Ph.D. New York: E. P. Dutton and Company. 1926. 217 pp.

The work of economists, statisticians and criminologists, all of whose results are reviewed in the early pages of Miss Thomas's work, had made it plain that there was an association between the different phases of the business cycle and such social phenomena as births, deaths, marriages, divorces and crimes. The studies of Denis, Tugan-Baronowsky, Farr, Ogle, Hooker, and Yule especially have been models for many less notable researches. A very distinctive study was published in 1922 by Professor Ogburn and Miss Thomas[1] in which was sought the correlation of the above social phenomena with a consolidated index of the business cycle, on the basis of American data. The present work by Miss Thomas, under the stimulus of Professor Bowley, parallels that study, by the use of English data.

Her results are, for the most part, what would have been expected in the light of previous study, but she has added some interesting discoveries and some anomalies. She finds that the association of marriage rate and business trends

[1] This JOURNAL, December, 1922.

has lessened in recent years, presumably because the increased use of contraceptives has enabled Eros to have his way with little danger of additions to the family. Although arrests for prostitution increase with prosperity and decrease with depression, this is believed to be due to the increased use of alcohol with prosperity, which may have the effect of actually concealing the true relation of prostitution and business conditions. Here is a beautiful illustration of the difficulties of social statistics. Arrests for prostitution, which constitute the only reliable measure of that institution in many communities, may actually vary inversely with the institution itself. One cannot be certain.

The anomalies appear in the apparent direct correlation of both the general death-rate and the infant death-rate with the rise and fall of business prosperity. The results for England are similar in this respect to those found by Ogburn and Thomas for the United States. If these correlations are correct the reasons therefore are not sufficiently evident. In the case of infant mortality Miss Thomas advances the theory that the correlation may be due to (a) the increased employment of women during prosperity leading to an increase in bottle feeding and infant neglect; (b) the increase in births resulting from the increase in marriages; and (c) the increased consumption of alcohol. But, as the author indicates, it would be presumed that increased prosperity would affect infant mortality favorably. The Galton Laboratory studies have purported to show that the employment of mothers, *taken by itself*, was a favorable factor presumably because it increased the family income and thus made possible foods and protection not otherwise attainable. They also showed that bottle feeding is an almost negligible factor when account is taken of the physique and intelligence of the mother. It may be that business prosperity increases disproportionately the number of mothers of very inferior type who are employed, mothers who succeed in keeping their infants alive if they remain at home, but who lack the energy or management to work away from home and also save their offspring. One wonders also if there may not be some direct conflict between the mother instinct and the desire to earn, a conflict which can only express itself when business prosperity opens the way for employment on good terms. Such conflict may well become serious where families are already large. It will require more investigation and the utilization of partial correlation methods to clarify this matter.

The same is true of the direct correlation of the general death-rate with the business cycle. As the author says, there are many factors at work in such a result; consequently little value can attach to so general a result as is here found. Moreover, she has eliminated all epidemic diseases, for what seem inadequate reasons; some of these are worthy of independent study. Nevertheless, Miss Thomas has made a valuable contribution to social statistics. Her methods have been carefully worked out and the whole study is worthy a high place in the literature of an exact quantitative approach to social analysis.

F. H. Hankins

Smith College

A Graphic Table Combining Logarithms and Anti-logarithms, by LaCroix and Ragout. The Macmillan Company. 1925. 64 pp.

Logarithm tables have become such staid and commonplace tools that an attempt to present logarithms in a different dress is refreshing, and the present book is welcome for that reason if for no other. The authors have taken a linear scale and graduated it with uniform divisions above the line and with divisions on a logarithmic scale below the line so that a number may be read above the line, and its logarithm below. The scale is about 365 feet long and is broken up to make forty pages of twenty-five lines each. To save space in the printing the authors found it convenient to change the scale at various points in the book with the result that numbers are not uniformly spaced throughout the book and this causes a slight loss of time in finding a particular number. The scale is so graduated that five place numbers and logarithms can be read directly without interpolation, and in this fact lies the advantage of the book.

The reviewer used the book in general computation for several months and came to the conclusion that in spite of the elimination of interpolation there was very little if any saving in time over the ordinary form of logarithm table. As a test of this question a list of one hundred, five figure numbers was chosen and their logarithms were looked up in Vega's seven place table and then in the graphic table. The seven place table was used in order to compare two different methods of avoiding interpolation. With the graphic table the time taken to look up one hundred logarithms was 29 minutes and 10 seconds, while with the Vega seven place table the time was 24 minutes. On the reverse operation the times were 31 minutes and 40 seconds for the graphic table, and 30 minutes and 50 seconds for Vega. Thus in this test the graphic table proved to be slightly slower than the regular table, and although further experience would probably speed up the use of the graphic table, it is doubtful if it would prove to be faster on routine computation than the ordinary form of table.

The last six pages of the book present a graphic table for four figure numbers and logarithms. In this table and in the five place one it is possible to read an extra figure with a satisfactory degree of accuracy.

The book has the advantage of compactness and this, together with the elimination of interpolation, makes it a desirable addition to the library of any computer.

LOWELL J. REED

Johns Hopkins University

Has Poverty Diminished? by A. L. Bowley and Margaret H. Hogg. London: P. S. King and Son, Ltd. 1925. viii, 236 pp.

Professor Bowley and Miss Hogg have presented the results of a study of the changed position of the poverty line in England which will arrest attention both on account of the significance of the findings and the methods employed. Poverty has diminished in at least four of the five towns in which investigations were made in 1924 and compared with earlier and similar investigations in the same towns made in 1913 and 1914.

The towns selected, with the exception of Stanley, a mining town near Durham, all give employment in more than one industry, though boots and shoes lead in Northampton, iron work in Warrington, biscuits and printing in Reading, and fine spinning in Bolton. A total of 4000 working-class families were selected by a process of random sampling and their economic position determined by applying the Rowntree minimum standards, slightly revised for 1924, to their incomes. The application was made both to full-time weekly wages and to the actual earnings in the week preceding the house visits.

With the single exception of the mining town, there was found to be a higher proportion of working-class families "certainly above" the minimum standard in 1924 than in the earlier years (95.3 per cent as over against 88 per cent) and smaller proportion "certainly below" (3.6 per cent as over against 11 per cent, full-time wages assumed in both periods). In the case of Stanley, a slight change in the other direction had occurred. These significant shiftings were found to have resulted from two factors. The first and most effective was the increase in real wages. Money wages had doubled in the interval while prices had risen only about 70 per cent. In families in which a man was normally earning, the proportion in poverty was found to be only one-fifth of that in 1913 (if full-time wages were assumed), but a little over one-half if "the maximum effect of unemployment is reckoned."

The other factor taken to explain the diminution of poverty was the lower birth-rate which had resulted in a smaller proportion of families with dependent children and consequently fewer needs to be satisfied. It was computed that the smaller families accounted for half as much of the improvement as that brought about by increased wages. The average number per family of children under 14 for all working-class families fell from 1.5 in the earlier period to 1.2 in 1924, a fall which was greater than that which occurred in all classes in England and Wales from 1911 to 1921, when the average for children under 14 dropped from 1.29 to 1.12 per family. It was apparent that the conventional family of five, mother, father, and three children, which Professor Douglas has demonstrated to be falsely assumed as representative, was not characteristic of this group. In fact, it was found to account for only 5 per cent of the working-class families.

Principally on account of the small increase in population, over-crowding (except in Stanley and Warrington) was not markedly greater than at the time of the former investigation. Over-crowding was measured by the proportion of families in which there were more than two persons per room. This standard, however, failed as a satisfactory basis of comparison because of the greater proportion of adults in 1924. Rents had risen about 50 per cent, but they took a smaller proportion of the income than in the earlier years. The housing accommodation in general seemed to the investigators "disproportioned to the income." The tenants, who had benefited by the Rent Restriction Acts, had developed no higher standard and were "unwilling to devote an adequate part of income to rent."

The present study does not follow closely the English tradition in investigations of this kind, which usually puts more emphasis than is customary in this coun-

try on the testimony of individuals called as witnesses rather than upon extended statistical measurement. This was the method used in the investigations in connection with the establishment of the wage boards in England and the earlier procedure in connection with the factory acts. The advantages of getting a picture in its human terms with all the realistic appeal which the individual witness makes has been more relied upon than the weight of cumulative experience. In the present study, on the contrary, the attempt has been to make a statistical presentation, but so much particularization is still held to that it is difficult and sometimes impossible to discover in the tabular form the central tendency which is its only excuse for being.

An instance in point is the ingenuous arrangement for packing facts into such tables as appear on page 172. It is hard to make out where and to what extent over-crowding begins from the arrangement which gives the number of houses in which a specified number of rooms were occupied by increasing numbers of "equivalent adults." One is not let into the secret of how the "percentage over-crowded" is derived. The authors have gone to great trouble to show the actual composition of every family visited. (Appendix). A first glance at some of the pages gives the impression that elaborate mathematical formulae are being employed but one discovers that the symbols, *mb* wd'_2 sc_1 *in l*, are given to inform the reader that in the group of families in which the earners were one man and one child, only one family was found in which that man was between 18 and 70 and that child earner a boy between 14 and 16 while the rest of the group consisted of a wife under 70 and two unmarried daughters living with parent, over 18 years, five boys or girls between 5 and 14, and one boy or girl under five. Having penetrated this labyrinth the reader is uncertain why he came. We are told that there were "more than 350 categories of families." A category, unfortunately, has only the significance of the basis on which it is made and the use that is made of it for purposes of classification. Some of the problems of preliminary classification continued to vex the authors to the end of the study. A troubled note is sounded again and again in the discussion of the housing situation. It seemed impossible to decide when a scullery is not a scullery and when a scullery is a room or not a room.

The reviewer wishes that the all-important effect of unemployment were more fully dealt with. This is not to be disposed of by a simple comparison of the actual earnings during the week preceding the investigation with the full-time wages. The economic position of the family must surely vary according to the length of the preceding unemployed interval. The authors speak confidently of reckoning "the maximum effect of unemployment." This reckoning is of greatest importance. It may be that the actual condition of the 4000 families was not as far above the poverty line as it has been assumed to be. In one instance the survey is spoken of as "too favorable" in this particular. However, the fact that the number of unemployed registered in all the towns was only 9.6 per cent may indicate that the towns studied were not as much affected by the prevailing unemployment as other parts of the country.

AMY HEWES

Mount Holyoke College

American Foreign Investments, by Robert W. Dunn. New York: Viking Press. 1926. 421 pp.

This work which was prepared and published under the auspices of the American Fund for Public Service, Inc., of New York City (supported by the Garland Foundation) is to be welcomed as the first extensive volume on the subject of our investments abroad. A considerable number of articles have been written and many speeches made and recorded on various aspects of the subject but until the appearance of Mr. Dunn's book there was nothing in print which gave a comprehensive view of the scope and variety of our foreign loans and investments.

In blazing a trail through a virtually unexplored field of economic and statistical study, Mr. Dunn has been eminently practical. Instead of discussing the economic aspects of the export of capital or of analyzing in detail the forces which brought about our shift from a debtor to a creditor position in international account, Mr. Dunn has devoted most of his efforts to presenting a practical summary and catalogue of our existing foreign loans and investments.

This volume is therefore not so much a study or a treatise as a work of reference to which the general reader will turn whenever he wishes to get some idea of the extent and character of the direct investments Americans may have made in any given foreign country. As such the volume will prove quite useful. The reader need not, however, expect a wealth of detail on specified investments in any country, for that implies the taking of an elaborate census for which we may have to wait long, although the Finance and Investment Division of the Department of Commerce has been considering an investigation of this character. But let it be said to Mr. Dunn's credit that he has shown great diligence and care in compiling and sifting his facts and figures from so many scattered sources; a glance at the numerous footnote references to sources shows that the author searched far and wide. Of course there are some typographical errors and there are statements whose present accuracy or completeness may be queried, but none of these is really serious.

The volume has been described as primarily a work of reference. The second half, however, may be regarded also as a book of source information, for it contains the text of a considerable number of bankers' loan contracts and concessions of various sorts. These contracts were collected and edited by Mr. Adrian Richt who has supplied brief introductory comments on the types of contracts selected for special treatment. It is to be noted, however, that to indicate the importance and significance of these undertakings Mr. Richt has relied more largely upon extensive italicizing of certain clauses and phrases in the agreements than upon his own direct comments. Mr. Richt's contribution to the volume is, to use the language of the author's foreword, for the purpose of giving "the careful reader some valuable exhibits in American economic penetration abroad." The reader must be "careful" indeed and also very well informed as to the actual circumstances under which these loans were made, if he is to have a correct understanding of the true significance of certain loans and concessions. It would have been helpful, for example, in the discussion of

American loans to Nicaragua, to which 46 pages are devoted, if the reader had been informed as to the seriously disturbed political and economic situation in Nicaragua at the time the first loan was made.

GROSVENOR JONES

Tables of the Probable Error of the Coefficient of Correlation as found by the Product Moment Method, by Karl Holzinger. Tracts for Computers. No. XII. London: Cambridge University Press. 1925. Quarto, 1–35 pp.

This is one of the Tracts for Computers, edited by Karl Pearson, and issued by the Department of Applied Statistics, University of London, University College. There are two pages of introduction and explanation of the use of the tables by the editor. Briefly, the tables are designed to avoid the labor of computing the function of two variables

$$.6744898\ (1-r^2)/\sqrt{N}.$$

This is the probable error of the coefficient of correlation r, determined by the product moment method, where N is the size of the population or universe under discussion. The formula is based on the assumption that r follows a normal distribution and the editor carefully points out under what conditions and for what sets of values of r and N it may safely be used. The tables begin with $N=20$, but for these small samples, running up to $N=50$, and $N=100$, the application of the formula for values of r beyond .04 and .06, respectively, is doubtful unless the distribution of r is known to be normal or very nearly so. With these limitations the tables may be found very helpful to those who do not have access to the usual types of computing machines. It is admitted that with a machine and Barlow's Tables at hand—and this is not the case with many workers who use correlation coefficients—these tables would probably not be required.

They are set up in two parts, Table A and Table B. The first gives the probable error for all values of N from 20 to 100 by units and the proportional parts tables necessary for the correction to r, which is given directly from $r = .00$ to 1.00 by intervals of .02. The values of the probable error appear on the right-hand page and the corresponding proportional parts on the left-hand page, making it necessary to open the book at one place only for a given calculation.

Table B proceeds from $N=100$ to $N=1000$ by intervals of 10, with r as before from .00 to 1.00 by intervals of .02. The variation of the probable error function in this table must be corrected for both N and r and so there are two proportional parts tables on the left-hand page to go with the values of the function given on the right-hand page.

The values of the probable error function are given to four decimal places and the proportional parts so arranged that by a simple calculation the interpolated value may usually be obtained to four decimal places, although the value is occasionally off one unit in the fourth place—a matter of no consequence.

The printing and general set-up is good, and one could quickly learn to use the tables and probably, with practice, find the required values quite as rapidly as with a machine.

The tables were prepared by Professor Karl Holzinger of the University of Chicago, and he is entitled to all due credit for lightening the burden of calculation for his fellow workers in the varied fields of statistics where correlation coefficients are used and their probable errors required.

JAMES W. GLOVER

University of Michigan

Coal. Economic Problems of the Mining, Marketing and Consumption of Anthracite and Soft Coal in the United States, by Edward T. Devine. Bloomington, Ill.: American Review Service Press. 1925. 448 pp.

The United States Coal Commission did an excellent job, as one of its members has said, " in digging coal facts" but not nearly enough "in sizing and preparing them for the market." Several books have already attempted to make good this deficiency, and of them the compact little volume on *Coal* by Dr. Edward T. Devine—himself a distinguished member of the Commission—may well prove useful to the largest group of readers. For it is far less formidable and far more readable than *What the Coal Commission Found* (by several of the Commission's experts), though it contains nothing as good as Mr. F. G. Tryon's masterly introduction to the latter; and the ordinary student will prefer Dr. Devine's tolerant simplicity of statement to the more biting analysis and the infuriating brilliance that Mr. Hamilton and Miss Wright have employed in *The Case of Bituminous Coal.*

"How the miners work and how their families live; how the operators manage their mines and what profits they make; how the public interest is looked after or overlooked,"—what the Commission learned about all these things is summarized admirably in Dr. Devine's pages, in the simplest language and with the aid of a hundred easily-interpreted graphs. First "The Characters" are led genially across the stage—"The Coal Itself, silent, inert;" "The Owners" sometimes "nearly as silent and inert;" the army of operators, straggling in the bituminous fields, marching in close order in anthracite, and the other leading actors; and finally "The Government," so far merely "a minor character," and "The Public" cutting the most absurd figure of all. Then follow useful summaries of the evidence in the case of each of the halves of the industry and in the case of the transportation of coal. These chapters are excellent, though Dr. Devine makes no claim to carry beyond other scholars the needed work of further analysis of the abundant data on wages and profits, and his geniality does not keep him from the statement of vigorous conclusions. For anthracite, "the aggregate profits are enormous . . . and they have been increasing very considerably." For the over-development of bituminous, "such is the unhealthy state of the industry that thus far this dropsical condition, instead of curing itself, gets worse from year to year."

The summaries, however, are very largely static in character; and at least for an industry that is changing as rapidly as the bituminous, this is a somewhat serious defect. To be sure, Dr. Devine discusses the growing consolidation movement and gives a few paragraphs to the possible effects of giant power, but

he pays only the slightest attention to the series of rapid changes in technology that are beginning to change the soft coal mines into factories. The omission of the effect on changing technique lends some unreality to the discussion of "The Worker's Right to Satisfaction in His Work," and it appears to lead to an over-statement of the power of the United Mine Workers and to an understatement of the extent of the problem of over-development. Perhaps it is significant that the engineers and technicians who appear as the *dei ex machina* of a happy ending in *Coal Age* and even in the comments of Cabinet members and in the reports of the Commission itself, and who are the smiling villains or at least the messengers of evil tidings in *The Case of Bituminous Coal*, are not even cast for a separate rôle in Dr. Devine's drama.

The author's virtue of terseness, moreover, deserts him in the final section on "A National Policy." This consists mainly of a series of homilies addressed to the various characters in the play. "The bituminous coal operators," for example, "can help matters by uniting unto larger units." Owners may "pay their taxes cheerfully." "The President might take the initiative." The miners are singled out to stop "telling smutty stories" and to "practice the simple virtues" which social workers expect of their clients, while most of the wholesalers (as well as two hundred thousand miners) are honored by a call to the more heroic virtue of "honorable self-elimination." "You and I," meanwhile, might "give more attention . . . to the furnace" and change our underclothes with the seasons!

But with these health hints is mixed much shrewd economic analysis, and the section is valuable for the crisp disposal of rival remedies with which it begins and for the vigorous defense of the Commission's remedies near its close. "Nationalization . . . is not a live issue" in the present state or "the national temper." "Trustification," in spite of the relative order it has brought to anthracite, "is not the way out. . . . The anti-trust laws reflect a public opinion which is not weakened but on the contrary enormously strengthened by the facts of existing combinations." "Finally, if there is any one thing above all others which the nation cannot afford to do, it is to do nothing." As against these policies, Dr. Devine offers the recommendations of the Commission—notably the Federal tax on royalties and differential profits in anthracite and a very considerable degree of inspection and regulation by the Interstate Commerce Commission—"as logical and consistent next steps." If anyone is inclined to criticize the author or his colleagues for the mildness of these remedies after the gravity of their diagnosis, he should at least recall that they are more radical than any that are likely to be adopted in the near future. But "if these recommendations are ignored now"—so Dr. Devine warns us—"they will have to be rediscovered and reformulated—perhaps in less favorable circumstances." It is a prophesy that Mr. Justice Sankey and Sir Herbert Samuel may well echo from across the Atlantic!

CARTER GOODRICH

University of Michigan

Educational Statistics, by C. W. Odell. New York: The Century Company. 1925. ix, 334 pp.

The book on review adds one more to the very numerous group of more or less elementary works on statistics, either general, or applied to specific subjects, which have been placed on the market by a variety of publishers in recent times. The purpose of the book can be best indicated by quoting the closing paragraph of the editor's introduction:

> This volume is designed therefore to meet two demands; first, the need for an elementary text in educational statistics for colleges of education and teachers' colleges; second, the need for a discussion of statistical procedure for the use of superintendents of schools and others whose professional training has not included work of this nature.

With this limited aim in view, the author has presented in simple and clear language the rudiments of statistics under the following chapter heads:

I. Tabulation and classification
II. Graphs
III. Averages
IV. Measures of variability
V. The relationship between two variables
VI. Unreliability
VII. The correlation and regression of three or more variables
VIII. The correlation of attributes
IX. Miscellaneous

There are four appendices, the first a bibliography, the other three tables relating to the probability integral and coefficients of correlation.

The presentation throughout is concise and simple and will probably be found to meet fairly well the requirements of the particular class of readers for whom the book is intended. Like many other books of an elementary character in most instances this work gives formulae without their derivation. To some readers this may be irritating, but the presumption is that readers of this class will turn to other works which will satisfy their wants but which would be beyond the limitations of persons to whom statistics is a tool for occasional use, rather than a subject for exhaustive study.

It is difficult to work up any enthusiasm for a guide to rule-of-thumb statistics, however excellent in its class it may be. But no doubt there is a distinct demand for manuals of this type from those who, for lack of time or other reason, desire to benefit from the results of thought without the trouble of thinking. For such, the book under review will no doubt be found serviceable. The presentation is on the whole simple, straightforward and as clear as it can be under the circumstances. The typography is exceptionally good.

ALFRED J. LOTKA

RECENT LITERATURE

STATISTICAL BOOKS

ANGELL, JAMES W. *Theory of International Prices.* Cambridge, Mass.: Harvard University Press. Harvard Economic Studies, Vol. 28. 1926. 585 pp.

BLANCHARD, RALPH H. *Workmen's Compensation in the United States.* International Labour Office. Studies and Reports, Series M (Social Insurance), No. 5. 1926. 103 pp.

CLARK, JOHN MAURICE. *Social Control of Business.* University of Chicago Press. 1926. xviii, 483 pp.

FOOD RESEARCH INSTITUTE. *The World Wheat Situation, December 1925 to March 1926.* Wheat Studies, Vol. II, No. 6. Stanford University, California. 33 pp.

INTERNATIONAL LABOUR OFFICE. *Wage Changes in Various Countries, 1914 to 1925.* Studies and Reports, Series D (Wages and Hours), No. 16. 1926. 143 pp.

MILLER, MARGARET S. *The Economic Development of Russia from 1905–1914 with Special Reference to Trade, Industry and Finance.* London: P. S. King and Son, Ltd. 1926. 325 pp.

NATIONAL INDUSTRIAL CONFERENCE BOARD. *The Agricultural Problem in the United States.* 1926. xiii, 157 pp.

C. F. TRENERRY. *The Origin and History of Insurance.* London: P. S. King and Son, Ltd. 1926. 344 pp.

GOVERNMENT PUBLICATIONS

UNITED STATES.

——Bureau of the Census.

"Financial Statistics of States, 1924."
"Mortality Statistics, 1923."

——Department of Commerce.

Bureau of Foreign and Domestic Commerce:

The Balance of International Payments of the United States in 1925, by Franklin W. Ryan, with a foreword by Herbert Hoover.

——Department of Labor.

Bureau of Labor Statistics:

No. 398. "Growth of Legal Aid Work in the United States."
No. 401. "Family Allowances: A Development of the Living-Wage Problem."
No. 402. "Collective Bargaining by Actors."
No. 404. "Union Scale of Wages and Hours of Labor, May 15, 1925."
No. 405. "Phosphorus Necrosis in the Manufacture of Fireworks and in the Preparation of Phosphorus."
No. 410. "Safety Code for Paper and Pulp Mills."
No. 411. "Proceedings of the Twelfth Annual Convention of the Association of Governmental Labor Officials of the United States and Canada."

——Treasury Department.

Public Health Reports:

Vol. 41.

No. 21. "Endemic Goiter and Intelligence of School Children."

No. 24. "Preliminary Birth, Death, and Infant Mortality Rates, 1925."

FOREIGN COUNTRIES.

Belgium.

"Statistique des Accidents du Travail, Exercice 1921, Exposé des Résultats." 59 pp.

"Rapport sur la Réparation des dommages Résultant des Accidents du Travail pendant les années 1921–1922–1923." 187 pp.

India.

"Report of the Committee of the Bengal Chamber of Commerce for 1925, Vol. I." xvi, 383 pp.

Japan.

"The Statistics of the Department of Commerce and Industry, 1924." 135 pp.

Switzerland.

"Mouvement de la Population de la Suisse pendant l'Année, 1923." 99 pp.

Western Australia.

"Statistical Register of Western Australia for 1924–25 and Previous Years."

ARTICLES IN PERIODICALS

MONTHLY LABOR REVIEW.

May, 1926. "Vacations with pay for wage earners"; "Family-allowance systems in foreign countries"; "International statistics of working population"; "Growth of legal aid work in the United States"; "Wages and hours in slaughtering and meat packing"; "Unemployment in foreign countries."

June, 1926. "Industrial Accident Prevention Conference, Washington, D. C., July 14, 15, and 16, 1926."

JOURNAL OF POLITICAL ECONOMY.

June, 1926. E. W. Kemmerer: "Chile Returns to the Gold Standard"; H. G. Brown: "Land Rent and Population Growth"; L. C. Marshall: "The Collegiate School of Business at Erehwon"; W. H. Steiner: "Rediscount at Federal Reserve Banks"; G. G. Benjamin: "Socialists and the Agrarian Question"; P. D. Converse: "Some Notes on Marketing Functions"; D. A. Pomeroy: "The Uniform Mortgage Act."

POLITICAL SCIENCE QUARTERLY.

June, 1926. W. Y. Elliott: "Mussolini, Prophet of the Pragmatic Era in Politics"; E. R. A. Seligman: "The Social Theory of Fiscal Science"; C. H. Woody: "Is the Senate Unrepresentative?"; F. Franklin: "Economic Theory and Economic Criticism—Cassel on Rent and on Interest"; A. F. Hinrichs: "The Cost of Tax-Exempt Securities"; J. R. Collins: "Karl Marx and Samuel Gompers."

THE ANNALS.

May, 1926. "Modern Crime: Its Prevention and Punishment."

July, 1926. "The United States in Relation to the European Situation."

JOURNAL OF ACCOUNTANCY.

May, 1926. L. L. Thwing: "Obsolescence"; Y. Watanabe: "Accountancy in

Japan"; L. L. Smail: "Simplified Treatment of Ordinary Annuities"; H. F. Boettler: "Significant Economic Development Since the Armistice."

June, 1926. "American Institute of Accountants Examinations, May, 1926"; H. T. Chamberlain: "A Discussion of Surplus with Reference to Surplus Available for Dividends"; J. L. Dohr: "The Board of Tax Appeals and Net Income."

July, 1926. A. F. Voke: "Accounting Methods of Colonial Merchants in Virginia"; W. M. LeClear: "Consolidated Balance Sheets and Contingent Liabilities"; G. Armistead: "Auditing Counties in Texas."

JOURNAL OF THE INSTITUTE OF BANKERS.

May, 1926. B. Campion: "Banking in Relation to Limited Companies, Gilbart Lecture II (Cont.), III, IV"; N. Crump: "The London and New York Markets in the Autumn of 1925."

June, 1926. "Annual Reports of Local Centers"; B. Campion: "Banking in Relation to Limited Companies, Gilbart Lecture IV (Cont.)."

JOURNAL OF THE ROYAL STATISTICAL SOCIETY.

March, 1926. R. J. Thompson: "The Productivity of British and Danish Farming"; H. C. Scott: "Some Administrative Aspects of the Rignano Scheme of Inheritance Taxation"; Editor of the "Statist": "Wholesale Prices of Commodities in 1925"; A. L. Bowley: "Precision of Index Numbers"; Major Greenwood: "Professor Tschuprow on the Theory of Correlation"; H. E. Soper: "The Moments of the Hypergeometric Series."

INTERNATIONAL LABOUR REVIEW.

April, 1926. P. Devinat: "The American Labour Movement and Scientific Management"; U. Ricci: "The Compilation of Wages Statistics"; "The New British Pensions Act: II"; "Industrial Inspection in Germany in 1923 and 1924"; "Industrial Diseases: Analysis of Factory Inspection Reports, 1920–1922: III."

May, 1926. M. Eastman: "A Brief Survey of Coal-Crisis Literature"; C. R. Clee: "Desertion and the Freedom of the Seamen"; H. M. Vernon: "The Human Factor and Industrial Accidents"; "The Problem of Labour Output in Soviet Russia"; "Child Labour in German Agriculture in the last Twenty Years."

June, 1926. I. Fisher: "A Statistical Relation Between Unemployment and Price Changes"; C. Biondi: "Medical Jurisprudence in Social Insurance and the Problem of Unification"; C. R. Clee: "Desertion and the Freedom of the Seamen (concluded)"; "Industrial Diseases: Analysis of Factory Inspection Reports, 1920–1922: IV"; "The Effects of Swedish Legislation on Hours of Work."

ZEITSCHRIFT FÜR VÖLKERPSYCHOLOGIE UND SOZIOLOGIE.

June, 1926. F. Schneersohn: "Die Kritik der Lehre von psychischer Infektion (resp. psychischer Epidemie) und die objektive Aneignungstheorie"; L. Benario: "Zur Soziologie der Zeitung"; A. Eleutheropulos: "Exakte oder spekulative Soziologie?"; Rehm: "Aus der kriminalpsychologischen Literatur"; P. Plaut: "Aubenseiter der Gesellschaft"; R. Woldt: "Die wirtschaftliche Organisationsumwelt des Arbeiters."

JOURNAL DE LA SOCIÉTÉ DE STATISTIQUE DE PARIS.

April, 1926. M. W. Methorst: "Méthodes à suivre pour la préparation des statistiques des stocks (fin)"; M. Yves-Guyot: "Prévisions rélatives aux paiements en nature des réparations et des dettes interalliées (fin)"; M. Edmond

Michel: "La situation financière et l'achèvement de la reconstitution des régions dévastées au 31 décembre 1925 (à suivre)"; M. Dugé de Bernonville: "Chronique des questions ouvrières et des assurances sur la vie."

May, 1926. M. Edmond Michel: "La Situation financière et l'achèvement de la reconstitution des régions dévastées au 31 décembre 1925 (suite)"; M. le Dr. Ichok: "Inflation et natalité"; M. Henri Bunle: "Chronique de démographie."

June, 1926. M. Edmond Michel: "La situation financière et l'achèvement de la reconstitution des régions dévastées au 31 décembre 1925 (à suivre)"; M. P. Ricard: "Chronique des banques et questions monétaires."

GIORNALE DEGLI ECONOMISTI.

April, 1926. Luigi Federici: "Appunti Critici sul Protezionismo Marittimo e su Alcune sue Forme Storiche"; G. C. Donvito: "L'Inflazione e il Mezzogiorno"; L. M. Severino: "Della Proprieta Privata."

May, 1926. F. P.: "Aspetti e Problemi dell'Economia Italiana: Le Industrie del Cotone in Italia"; G. Zingali: "In Atto del Censimento del Bestiame"; C. E. Bonferroni: "Sulle Medie di Potenze"; G. Angrisani: "A Proposito di un Saggio (Per L'Evento della Maternita)."

GIORNALE DI MATEMATICA FINANCIARIA.

February, 1926. F. Zaula: "Metodo dei gruppi e sistema di controllo per il conteggio delle riserve di premio delle assicurazioni vita"; E. Lenzi: "Sulle rendite, le costituzioni di capitali e gli ammortamenti in generale"; F. Insolera: "Sei anni di esperienza dell'assicurazione obbligatoria contro gli infortuni in agricoltura."

MISCELLANEOUS REPRINTS AND PAMPHLETS

RESIDENCE CONSTRUCTION AND OTHER FACTORS WHICH HAVE DETERMINED RENT LEVELS IN PITTSBURGH, PENNSYLVANIA. University of Pittsburgh Studies in Business Administration, Publication No. 3.

INDEXING OF FEDERAL AND STATE LAWS. Hearing before the Committee on the Judiciary House of Representatives, Sixty-Ninth Congress, First Session on H. R. 9173, 9174. Serial No. 9. April 2, 1926.

EXPENSES OF LIFE-INSURANCE COMPANIES OPERATING IN NEBRASKA. University of Nebraska Studies in Business, No. 15.

THE INFLUENCE OF EMPLOYMENT UPON THE FERTILITY OF WOMEN. Statistical Bulletin, Vol. VII, No. 5. Metropolitan Life Insurance Company.

THE VALUE OF A MAN AS A WAGE-EARNER. Statistical Bulletin, Vol. VII, No. 6. Metropolitan Life Insurance Company.

FORSCHUNGEN ZUR VÖLKERPSYCHOLOGIE UND SOZIOLOGIE, PARTEI UND KLASSE IM LEBENSPROZESS DER GESELLSCHAFT. G. Roffenstein: "Das Problem der Ideologie in der materialistichen Geschichtsauffassung und das Moderne Parteiwesen"; F. Giovanoli: "Zur Soziologie des Parteiwesens und Betrachtungen zur schweizerischen Demokratie"; C. Cornelissen: "Theoretische und ökonomische Grundlagen des Syndikalismus"; R. Kobatsch: "Mittel zur Milderung der Klassengegensätze"; F. Bohn: "Die psychologischen Wurzeln einer Parteibewegung in Amerika."

I PROBLEMI TECNICI DELLA STABILIZZAZIONE MONETARIA SECONDO L'ESPERIENZA TEDESCA, by Giovanni Lasorsa. Edizioni di *Economia*.

OTHER BOOKS

CHIAO, WAN-HSUAN. *Devolution in Great Britain.* New York: Longmans, Green. 1926. 287 pp.

FERNANDEZ, L. H. *The Philippine Republic.* New York: Longmans, Green. 1926. 202 pp.

GAZLEY, J. G. *American Opinion of German Unification, 1848–1871.* Longmans, Green. 1926. 582 pp.

KO, TING TSZ. *Governmental Methods of Adjusting Labor Disputes in North America and Australasia.* New York: Longmans, Green. 1926. 219 pp.

SULLIVAN, J. W. N. *Three Men Discuss Relativity.* New York: Alfred A. Knopf. 1926. viii, 249 pp.